

基于生成式人工智能的小学数学应用题学习环境设计及其研究

Design and Research of a Generative AI-Based Learning Environment for Primary School

Mathematics Word Problems

刘靖熙¹, 孙丹儿^{1*}, 郑之姿¹, 许志文¹

¹ 香港教育大学数学与资讯科技系

* dsun@eduhk.hk

【摘要】 研究设计了名为“ChatGPT支持的数学应用题解决系统”(ChatGPT-MPS),旨在帮助小学生提升应用题解决能力、加深数学概念理解,并促进基于AI的协作探究学习。研究采用随机对照试验设计,选取了104名五年级学生,并按班级分配至实验组(使用ChatGPT-MPS)和对照组(传统教学法)。通过前测和后测数据分析,结果显示实验组学生在数学问题解决任务中的表现显著优于对照组,同时其学习兴趣和数学学习的感知价值显著提升。研究结论表明,ChatGPT-MPS在增强学生的数学问题解决能力和学习积极性方面具有显著效果,为未来AI技术在教育中的应用提供了重要的启示和实践依据。

【关键词】 人工智能; ChatGPT; 小学; 数学; 应用题解决

Abstract: This study explores the potential of ChatGPT technology in primary mathematics education, addressing existing research gaps and highlighting its transformative capabilities in technology-enhanced learning. A system named the "ChatGPT-Supported Mathematics Problem-Solving System" (ChatGPT-MPS) was developed to enhance students' abilities in solving word problems, deepen their understanding of mathematical concepts, and promote AI-facilitated collaborative inquiry. A controlled trial was conducted with 104 fifth-grade students, who were assigned to either an experimental group (using ChatGPT-MPS) or a control group (traditional teaching methods) by their classes. Pre-test and post-test analyses revealed that the experimental group outperformed the control group in problem-solving tasks, with significant improvements in learning interest and perceived value of mathematics education. The findings demonstrate the effectiveness of ChatGPT-MPS in enhancing students' mathematical problem-solving skills and fostering greater enthusiasm for learning, offering valuable insights and practical implications for the integration of AI technologies in education.

Keywords: Artificial Intelligence, ChatGPT, Primary school, Mathematics, Word Problem solving

1. 引言

根据 Baker 和 Smith (2019) 的定义,人工智能(AI)指能够使计算机执行诸如学习和问题解决等认知功能的技术。随着 GPT 和 BERT 等大型预训练模型的发展,AI能力特别是在自然语言处理(NLP)任务(如预训练、微调和文本生成)方面显著提升(Min等,2021)。这些模型已在教育等多个领域展现出变革性作用。尽管AI在教育中的应用迅速发展,但在解决数学应用题这一结合语言理解和数学推理的特定领域中的研究仍然稀缺。现有研究多集中于提升整体学习体验,却很少探讨AI如何应对这一具体学习挑战,尤其是在教师支持有限的情况下。为填补这一研究空白,本研究开发并评估了“ChatGPT支持的数学应用题解决系统”(ChatGPT-MPS),创建一个基于人工智能的数学学习环境,旨在提升小学生解决数学应用题的能力。

2. 文献综述

人工智能(AI)以其强大的数据处理和实时反馈能力,在数学教育中展现了显著潜力。AI驱动的个性化学习系统能够根据学生的水平和需求,提供定制化的学习路径和内容,帮助学生掌握复杂的数学概念。数学学习的难点在于需要结合语言理解、逻辑推理和数学运算能力,但许多学生因阅读理解困难、运算技能不足以及数学焦虑等因素而面临挑战(Boonen等,2016; Doz等,2023)。AI通过分析学生的学习行为,识别知识薄弱点并提供即时反馈,有效解决了这些问题。尽管AI在数学教育中的应用取得了显著成果,现有系统仍存在局限,如缺乏实时互动和适应性支持(Bulathwela等,2022)。为弥补这些不足,本研究开发了“ChatGPT支持的数学应用题解决系统”(ChatGPT-MPS),专注于小学数学应用题的个性化解决。通

过结合 ChatGPT 技术与小学课程，系统提供实时反馈和动态支持，帮助学生提升数学理解与解题能力。

3. 研究目的

本研究旨在探讨 ChatGPT-MPS 对学生解决数学应用题能力的影响。具体而言，研究将评估使用 ChatGPT-MPS 的学生在解题能力上的表现，并与传统教学方法的效果进行比较，以揭示其在提升数学应用题解题能力方面的优势和潜力。

4. ChatGPT-MPS 系统架构

ChatGPT-MPS 系统专为小学数学教育设计，旨在利用 ChatGPT 的自然语言处理能力为学生提供个性化的数学问题解决支持。通过集成 ChatGPT API，该系统为学生创建了一个友好的学习环境，提供实时互动、个性化解题建议和多样化的解题路径。系统采用前后端分离架构，支持多达三种不同的解题方法，通过动态调整响应满足学生的多样化需求。

ChatGPT-MPS 通过 ChatGPT API 实现实时交互，支持学生以自然语言输入数学问题。系统在接收问题后，结合小学五年级的知识水平和特定约束条件生成结构化的逐步解答，同时存储多种解决方案以供比较与优化。通过多次点击，学生可以获取多样化的解题路径，并保存学习记录，用于后续参考和改进。

5. 研究方法

5.1 参与者

本研究涉及来自中国广东省某公立小学的两个班级的 104 名五年级学生。一个班级（52 名学生）作为实验组（使用 ChatGPT-MPS），另一个班级（52 名学生）作为对照组（使用传统教学方法）。两个班级的学业表现相当，之前的通过率约为 80%，成绩率在 23.1% 到 54% 之间。

5.2 程序

实验过程分为两个主要任务，分别为实验组和对照组。实验组学生使用 ChatGPT-MPS 系统完成 12 个数学应用题。任务开始前，学生参加 10 分钟的教学环节，了解系统功能和操作指南。随后，学生在系统指导下用 40 分钟完成题目。任务结束后，随机选取 20 名学生参与可用性测试。此外，还对监考人员和观察教师进行访谈。

对照组完成相同的 12 道数学应用题，但未使用 ChatGPT-MPS 系统，而是依靠前序知识、课程知识和解题能力独立完成。任务同样限时 40 分钟，试卷收集后随机选取 20 名学生进行可用性测试，评估其对任务清晰度和难度的感知。通过对两组学生的表现、行为数据和反馈的对比，研究旨在评估 ChatGPT-MPS 系统在提升学生数学解题能力和任务可用性方面的实际效果。

5.3 数据收集

前测和后测经过精心设计，包含相同的测试题目，以确保对学生使用 ChatGPT-MPS 系统前后解题能力的有效且可靠的比较。这两项评估各包括 12 道符合五年级课程要求的数学应用题。测试的有效性通过多重手段加以保障。首先，内容效度通过严格遵循 2022 年义务教育课程计划和课程标准建立，确保测试题目准确反映五年级的数学学习目标。

5.4 数据分析

5.4.1 数据清理

在进行统计分析之前，实施了彻底的数据清理过程。如图 1 和图 2 所示，两种情况下分数的分布均表现出集中趋势，大多数分数集中在中间范围，从而满足了进一步分析所需的正态性条件。

图 1
任务后分数直方图

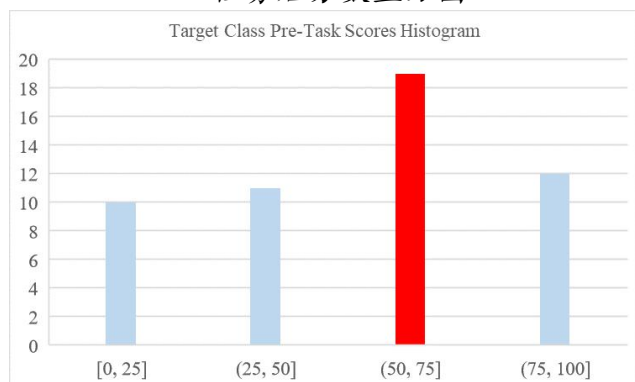
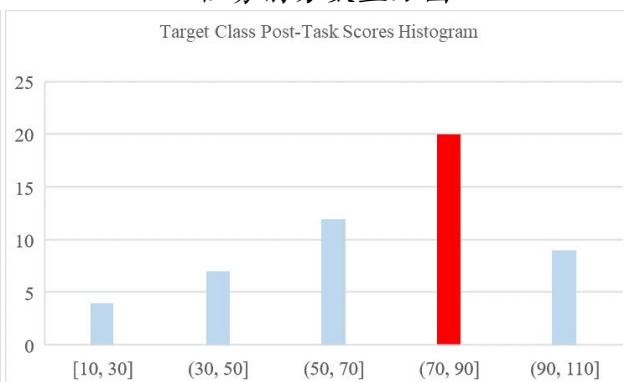
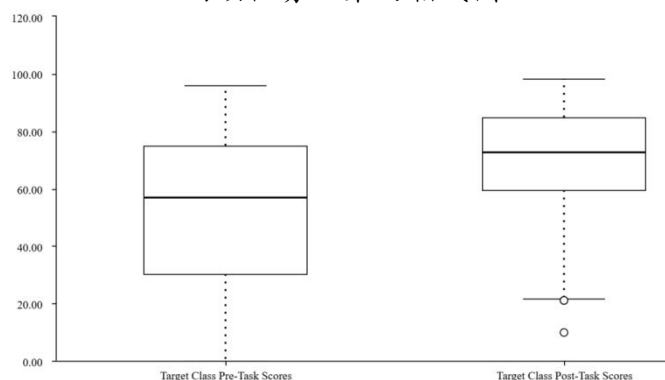


图 2
任务前分数直方图



随后，进行了箱线图分析（图 3），以识别潜在的异常值。在任务后的数据集中，观察到分数为 10、21 和 21 的异常值，这些异常值可能会对结果产生不当影响。为了保持分析的稳健性，排除了这些异常值，将样本量减少到 49 名学生。这一步骤确保了后续的统计测试能够提供可靠且无偏的结果，反映学生的实际表现。

图 3
两项任务结果的箱线图



5.4.2. 数学应用题解决表现分析

为了评估 ChatGPT-MPS 系统对学生解决数学应用题表现的影响，采用了配对样本 T 检验。该检验非常适合比较同一组学生在两种不同条件下的平均分数——干预前和干预后，从而控制个体差异（Ross 等人，2017）。选择此检验方法是为了确定使用 ChatGPT-MPS 系统是否导致学生解题能力的统计显著提高。配对样本 T 检验比较了任务前的分数（反映没有系统时的学生表现）和任务后的分数（使用 ChatGPT-MPS 后收集）。经过数据清理和异常值去除后，T 检验显示任务后的分数显著增加，表明学生的数学解题技能因 ChatGPT-MPS 系统的使用而得到了有意义的提升。

6. 研究发现

任务前的平均分数为 56.45，标准差为 24.37，显示学生初始表现存在较大差异。使用 ChatGPT-MPS 系统后，任务后的平均分数提升至 72.59，且标准差减少到 17.45，表明分数分布更加集中在较高的平均值周围。两个任务分数之间的相关系数为 0.763，显示出任务前与任务后得分之间存在强正相关关系。平均差为 -16.14286 分，说明学生在任务后平均提高了 16.14 分。标准差 15.79 和均值标准误差 2.26 反映了这一平均差的变化性和精确度。t 值为 -7.156，自由度为 48，表明任务前与任务后得分之间存在显著差异。数据确认了 ChatGPT-MPS 对学生解决数学应用题表现的正面影响。Cohen's d 值为 -1.022，Hedges' g 值为 -1.014，表明效应规模大，意味着 ChatGPT-MPS 对提高学生分数产生了重大影响。综上所述，配对样本 T 检验结果显示，在使用 ChatGPT-MPS 系统后，学生在解决数学应用题上的表现有了显著提高。大的效应规模表明该系统对学生数学能力的提升有实质性的影响，展示了 AI 工具在教育环境中使用的潜力。

7. 结论与讨论

本研究旨在探讨 ChatGPT-MPS 在提升小学生解决数学应用题表现中的作用, 重点分析其开发、测试过程及对解题能力的影响。使用 ChatGPT-MPS 后学生成绩显著提升, 任务前平均分从 56.45 分增至任务后 72.59 分, 平均提高 16.14 分。前后得分之间的强相关系数进一步证实了这一提升。个性化反馈是这一成效的关键机制, 帮助学生识别并解决具体的弱点。此外, ChatGPT-MPS 的互动式学习环境促进了学生对解题策略的深入理解, 实时的 AI 辅助进一步澄清误解并引导学生解决复杂问题。

本研究为将 AI 工具整合至教育环境提供了宝贵见解, 证明 ChatGPT-MPS 在提升学生数学解题能力方面的有效性。这些成果不仅展现了 AI 技术在教育中的潜力, 还为进一步优化教学工具设计和教育政策提供了依据, 为学生创造更高效的学习体验奠定了基础。

参考文献

- Baker, T., & Smith, L. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Boonen, A., Koning, B., Jolles, J., & Schoot, M. (2016). Word Problem Solving in Contemporary Math Education: A Plea for Reading Comprehension Skills Training. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00191>.
- Bulathwela, S., Pérez-Ortiz, M., Yilmaz, E., & Shawe-Taylor, J. (2022). Power to the Learner: Towards Human-Intuitive and Integrative Recommendations with Open Educational Resources. *Sustainability*. <https://doi.org/10.3390/su141811682>.
- Min, B., Ross, H., Sulem, E., Veyseh, A., Nguyen, T., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2021). Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. *ACM Computing Surveys*, 56, 1 - 40. <https://doi.org/10.1145/3605943>.
- Ross, A., Willson, V. L., Ross, A., & Willson, V. L. (2017). Paired samples T-test. *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures*, 17-19.