240517 Review of Research on classroom dialogue among different types of teachers based on large language models

**General comments**

Overall, this is an interesting study that used LLMs to investigate the differences in classroom dialogue between pre-service and experienced teachers. It is presented in a well-written paper that is easy to follow.

The paper might be of interest to journals focused on mathematics education and the training of mathematics teachers, such as:

[Journal for Research in Mathematics Education](https://pubs.nctm.org/view/journals/jrme/jrme-overview.xml)

[Journal of Mathematics Education](https://www.educationforatoz.com/)

[Journal of Mathematics Teacher Education](https://link.springer.com/journal/10857)

The only challenge you might have in submitting the paper to those journals is that your analytic methods would probably be unfamiliar to the readership of the journals, as the methods are quite technical. However, best to just submit and see what feedback you get. You could always simplify the discussion of the technical methods if needed.

The paper might also be of interest in more methodologically focused journals such as

[Journal of Educational Data Mining](https://jedm.educationaldatamining.org/index.php/JEDM)

Take a look through the paper to make sure that you are consistent in the use of either present tense (eg “the study shows….”) or past tense (eg “ the study showed…”). I think it might be best to use past tense throughout. Some journal papers are also written in the first person plural (eg, “We applied large language models to investigate….” Rather than “the study applied….”). Maybe you could take a look at the journals you are planning to submit to and see what their style is.

**My thoughts and suggestions**

Introduction

It is good that the research questions are stated at the end of the introduction. I think you could be clearer by changing the sentence that begins “In the study process…” to say “The study sought to answer the following research questions (1)…..”

The two questions you have are good, but maybe you could add a third question, something like…(3) How can pre-service teachers improve their classroom dialogue. I think the study can answer this question (or something like it).

However, although you identify the research questions here, in the results section you need to link the results to the research questions. At the moment, I don’t think you mention them after the Introduction.

2. Literature Review

The Lit Review is very good and is easy to follow. Well done!

2.1 Classroom Dialogue

Should there be a citation for the Cambridge University Classroom Dialogue Group?

2.2 Applications of LLMs in classroom dialogue

The end of the first sentence “and their proposal is of epoch-making significance” didn’t quite make sense to me. Perhaps just omit that part or re-write it.

2.3 Pre-service and expert teachers

A thought about the use of the term ‘expert teachers’ throughout the paper. There has been a lot of prior research in the field of teaching to identify truly ‘expert’ teachers – those who are true masters of their craft. In the USA there is even a qualification that teachers can seek to recognise that they are ‘accomplished’ (see <https://www.nbpts.org/certification/>). Although I see that you have a definition of what you mean by ‘expert teacher’, your use of the word ‘expert teacher’ might be confusing. I think what you mean is ‘experienced teacher’ – a teacher who is not a beginning teacher and has some years of experience in the profession. I don’t feel strongly about this if you decide to stick with your use of ‘expert teacher’ but I thought I would point it out.

Also, I had a thought about the research you cite on the differences between pre-service and experienced teachers. This research is, I presume, based on studies of Western classrooms. Your study is in Chinese classrooms, and I wondered if the differences that are identified in the Western studies also apply to the Chinese teachers. Given that your study did identify the differences in dialogue that you found, I presume that the differences are similar in both Western and Chinese classrooms, but you might want to say something about this if you can.

3. Methodology

Overall the methodology section is very clear and well written

3.1 Data Sources

I wasn’t clear if the language of instruction in the classrooms in your study was Mandarin. Or were students taught in English? I think you should say. Also, if it was Mandarin, did this have any implications for what LLMs you could use?

3.2 Data Processing

There is a typo on ‘vioce’ – should be ‘voice’

3.3

You should say what the abbreviation GSEQ stands for – Generalized Sequential Querier, I think.

4 Results

Overall, the results section is clearly written and easy to follow. The results are also interesting.

As mentioned earlier, though, you should link the results to the Research Questions you stated in the Introduction. You should say something like, “In answer Research Question 1, which asked ……, we found that the automatic coding abilites of the large language models we used were able to clearly identify patterns in the classroom dialogues and to see differneces between the dialogues of pre-service teachers and those of experienced teachers.” Then go on to explain in more detail the results as you have done.

4.1 Automatic coding results of classroom dialogue

I think that it would be easier for readers if you somehow make it clear when the words you are using are the names of the coding categories. For example you could use italics (eg, *Speculation* and *inviting coordination*), or you could use quotations (eg, ‘Speculation’ and ‘inviting coordination’).

In the sentence that begins, “Speculation and inviting coordination….” You say these are the top **three** categories, but I think there are only two mentioned.

To make it easier for the readers you could move Table 1 to the end of section 4.1 and then it would be easier for them to refer to it.

4.2 Situational analysis…

The sentence at the bottom of page 7 that begins, “Precision measures …” doesn’t quite make sense. Can you re-write it?

A general comment about Table 1 and all the other tables that use abbreviations of the coding categories. I know that there is a very clear table in Appendix A that explains all the coding categories and shows their abbreviations, but you don’t mention this in the results section, I think. It would be good to do so. Even then, I think it is a lot to expect of readers to be able to know what the abbreviations are in Tables 1, 2, 4, 5, 6 and 7, and Figure 2. It would be better if you could either use the full names in the first column of each table, or put a key at the bottom of the table.

Also, I think your table numbering has got out of order, as you have table 3 appearing before Table 2

Figure 2 – I found the text and numbers too small to read easily. Are you able to make it bigger?

Table 3 – I think it would avoid confusion if you add the word “utterances” to “Average number of students’ ” and “Average number of teachers’ “

4.4 Analysis of the evolution…

Figures 3 and 4 are hard to read. Can you make them bigger?

5. Discussion

As in the Results section, you can link the discussion to the original research questions.

Some final thoughts

Could you use the results to give pre-service teachers on optimal dialogue sequences that they might use? For example, you say on page 15 that the classroom dialogue sequences tend to be longer and more complex and you identify two particular paths. Would it be possible to extract some dialogue examples to illustrate what this looks like in practice? If so, you could annotate the sequences to show how experienced teachers guide the dialogue to achieve higher order thinking etc. Maybe, a further study?