



Peer-Review Report Form

Manuscript Title:	Student Perceptions and Learning Outcomes of ChatGPT in German Language Learning at Nangarhar University
Type of Article:	Original Article

FIRST ROUND

Editor

<u>General Comments and Recommendations</u>
The reviewers acknowledge the significance of your manuscript and recommend several revisions to further improve its clarity, coherence, organization, and overall scholarly quality. In addition, you are advised to thoroughly review and validate all references included in the manuscript. Please ensure that all cited sources are accurate, accessible, and retrievable through Google Scholar to uphold the credibility and reliability of the study. Additionally, make sure to include the DOI or website link for every reference to facilitate verification and accessibility of sources. Furthermore, only sources that are cited within the manuscript should appear in the References section, and all entries listed in the References section must also be cited in the text. For example, if 20 unique sources are cited in the manuscript, then only those 20 sources should be included in the References section. Please note that repeated citations of the same source within the text are counted as a single reference entry; therefore, the total number of references may be lower than the total number of in-text citations.

Reviewer 1

<u>General Comments and Recommendations</u>
The manuscript discusses a timely and relevant topic, especially in the Afghan higher education context where language learning resources may be limited. The study has value because it gives initial evidence on how German language students use ChatGPT. However, the paper still needs stronger methodological clarity and deeper analysis before its conclusions about learning outcomes and effectiveness can be fully accepted.

<u>Specific Comments and Recommendations</u>
The title and purpose mention "learning outcomes," but the study mainly measures student perceptions and self-reported use. There is no direct evidence of actual learning outcomes, such as test scores, writing samples, pre-test and post-test results, or language proficiency changes. This is important because positive student opinion does not automatically mean improved learning.
The methodology needs to explain the sampling process more clearly. The manuscript states that the population was 120 students and that 90 students were included, with three non-responses, but it is not clear whether the final sample was 90 or 87. The use of Yamane's



formula is also mentioned, but the calculation and final justification are not clearly presented. This affects the credibility of the sample.
The questionnaire is central to the study, but there is no clear information about its validation, reliability, pilot testing, or how the items were developed. Since the findings depend fully on student responses, the authors need to show that the tool measured perceptions, use, and impact in a dependable way.
The analysis is too descriptive for the claims being made. Frequencies, percentages, mode, and range can describe student responses, but they cannot strongly prove effectiveness or impact. The study would be stronger if it examined relationships, for example between frequency of ChatGPT use and perceived skill improvement, or between year level and AI familiarity.
There is an important inconsistency in the interpretation of incorrect information. Table 8 shows that 38.9% reported encountering incorrect information, while 61.1% said no. However, the discussion and conclusion state that about 61% experienced misinformation. This is a serious issue because it changes the meaning of one of the main findings.
Some tables and interpretations appear repetitive and not fully connected to the research problem. For example, several paragraphs only restate the table values without explaining what they mean for German language learning. The discussion should move beyond description and explain why writing and reading are more supported than speaking and listening.
The literature review includes many theories and studies, but the theoretical framework is not strongly used in the analysis. Vygotsky, CLT, and the Input–Interaction–Output model are mentioned, but the results are not interpreted using these theories. The authors should either connect these theories clearly to the findings or reduce the theoretical discussion to avoid appearing unfocused.

Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.

Reviewer 2

<u>General Comments and Recommendations</u>
I read your paper about the use of ChatGPT among German language learners at Nangarhar University with big interest. The topic you chose is extremely relevant for education today, and it is especially important for the context of Afghanistan where modern technology and educational resources are very limited. You did a solid job gathering data from 90 students to understand their feelings and experiences.
<u>Specific Comments and Recommendations</u>
Title and Abstract



1. Your title includes the phrase learning outcomes. This is slightly misleading for the readers. Because you used a survey questionnaire to collect opinions, you are studying perceived learning outcomes. You did not conduct exams or tests to measure actual German language proficiency. I strongly suggest you add the word perceived to your title.

2. In the abstract, you state that ChatGPT has become an effective source of instant feedback and skill development. It is much safer to write that students report it as an effective source, to stay true to your survey methodology.

Introduction and Literature Review

1. I really like how you describe the background of Afghanistan. You clearly explain the obstacles like the shortage of professional teachers and lack of practice opportunities. This gives your study very strong justification.

2. In the literature review, you introduce very excellent theoretical frameworks, specifically Vygotsky theory of social constructivism and the Communicative Language Teaching theory. This shows good academic thinking. But sadly, you forget about these theories completely in your Results and Discussion section. To make your paper deep, you must connect your survey findings back to these theories. For example, how exactly does the AI act as the supportive partner that Vygotsky talks about based on the students answers?

Methodology

1. There is a clear mathematical contradiction in your sampling that you must address. You state that you used Yamane formula to calculate the sample size. Your calculation shows the minimum required sample is 92.31 students. However, you mention you only collected data from 90 students because three did not respond. This means your sample size is slightly below the scientific requirement. You must acknowledge this as a limitation in your conclusion section.

2. The equations for the Yamane formula are typed very badly in the text and are difficult to read. Please format these mathematical equations properly using standard equation formatting so they look academic.

3. You describe your study as a descriptive-analytical quantitative design. But looking at your analysis criteria, you only used frequency, percentage, mode, and range. This is purely descriptive statistics. You did not do any analytical statistics like correlation tests to see if age or study year affects how often they use ChatGPT. I recommend you either remove the word analytical or run some cross-tabulations in Excel to show deeper relationships between your variables.

Results and Discussion

1. You have 14 separate tables for every single simple question. This takes up too much unnecessary space and interrupts the reading flow. Please combine all demographic data into one table, and all usage pattern data into a second table.

2. Your use of the statistical term Spread in the tables is not correct. For example, in Table 4, you write the spread is Low-Medium-High usage , and in Table 5 you write Broad purposes. This is not how variability is measured in statistics for categorical data. If you are just doing simple descriptive surveys, just stick to reporting the frequency and the percentage.

3. It is a very interesting finding that 61.1 percent of students say they have never encountered incorrect information , while 38.9 percent did. You should use critical thinking to discuss this more. Is it possible that the students who did not see wrong information just have a lower level of German and cannot catch the mistakes the AI is making? Exploring this will make your discussion much stronger.

Conclusion

1. Your conclusion summarizes the paper nicely and highlights that students use the tool to solve homework and practice skills.



2. You did a very good job identifying future research gaps, like looking at other languages or gathering qualitative data.
3. Because 88.9 percent of your participants want a training workshop, you should provide very practical recommendations at the end. What exactly should the administration at Nangarhar University do next month to help these students? Giving concrete advice will increase the practical value of your research.

Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.

SECOND ROUND

<u>Decision</u>	<u>Editor</u>	<u>Reviewer 1</u>	<u>Reviewer 2</u>
Accept the manuscript for publication.	/	/	/
Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.			
Reject the manuscript, as it is not suitable for publication.			