

Idea Sharing: Using ChatGPT for Diagnostic Writing Assessment

Napat Jitpaisarnwattana

Silpakorn University, Bangkok, Thailand

Corresponding author: napat.jit@gmail.com

Article information

APA citation:	Jitpaisarnwattana, N. (2024). Idea sharing: Using ChatGPT for diagnostic writing assessment. <i>PASAA</i> , 69, 561–574.
----------------------	--

1. Introduction

The emergence of generative AI, especially ChatGPT in 2022, has attracted much attention from many language teaching professionals and sparked a discussion regarding its technological, pedagogical, and ethical implications (Kohnke et al., 2023; Ulla et al., 2023). In a broad sense, the affordances and pedagogical benefits of ChatGPT hold the potential to create a paradigm shift in the language teaching and learning landscape (Jeon & Lee, 2023). Kohnke et al. (2023) outline several affordances of ChatGPT for both language learners and teachers. It offers learners opportunities for authentic interaction, helps them learn the meaning of words in context, and corrects and explains language errors. Findings from empirical studies also show that ChatGPT is useful for improving writing skills including grammar, vocabulary, coherence, argumentation depth (Nguyen, 2024), and formal register knowledge (Özçelik & Ekşi, 2024). As for teachers, they can develop quizzes, annotate texts, create lists of suitable vocabulary for each level, and even receive assistance with grading students' essays. Although using ChatGPT for summative assessments, such as grading, still seems impractical or even unethical, it offers potential for other forms of assessment, such as diagnostic or learning-oriented assessment.

It is a common practice for some language teachers to conduct pre-tests or diagnostic tests at the beginning of the courses they teach, either to simply understand the current level of their students or for diagnostic purposes. In my many years of teaching English paragraph and essay writing to undergraduate and graduate students, I often started my classes by asking them to write about a given topic, with the hope of using their writing to identify their weaknesses and, in some cases, altering my syllabus to fit their profiles. This is a practice shared by many language teachers who find knowing students' weaknesses useful for their syllabus design (Al-Mukdad, 2019; Nurbayan et al., 2020). However, it is such a laborious process that many often end up having to provide generic and/or group feedback instead. This is one of the main problems regarding feedback provision (especially formative/diagnostic feedback). The other problem is the immediacy of feedback. Many students often have to wait five to seven days or more to receive feedback from their teachers (Jitpaisarnwattana & Saville, 2024). One of ChatGPT's affordances is to identify and explain language errors, as well as suggest improvements immediately (Kohnke et al., 2023; Ulla et al., 2023). I believe that ChatGPT can serve as an effective assistant, providing detailed diagnostic assessments of students and offering language teachers valuable diagnostic data to use for feedback provision and course design.

This ideas-sharing article discusses the concept of diagnostic language assessment and showcases a step-by-step procedure for harnessing ChatGPT for diagnostic purposes. The article concludes with my reflective experience and offers some pedagogical and assessment implications for ChatGPT.

2. Diagnostic Language Assessment

In the last decade, diagnostic language assessment has garnered interest from several parties in language education, including language testers, language teachers, applied linguists, and, particularly, writing teachers and researchers (Alderson et al., 2015; Kunnan & Jang, 2011). This increased interest stems partly from a paradigm shift in language testing, from summative and product-oriented

testing to more formative and learning-oriented assessment. This has resulted in greater emphasis being put on identifying students' problems in language learning (and use) and finding effective ways for teachers and learners to deal with those problems (Lee, 2015). While summative assessment focuses on the performance at the end of a given course and formative assessment on progress during the course, diagnostic assessment is usually carried out at the beginning of the course to gauge what students know about a topic. Diagnostic language assessment (DLA) is defined as "the processes of identifying test-takers' (or learners') weaknesses, as well as their strengths, in a targeted domain of linguistic and communicative competence and providing specific diagnostic feedback and (guidance for) remedial learning" (Lee, 2015, p. 303). Despite DLA's primary focus on pinpointing weaknesses, it is interesting that, in this definition, strengths are identified in conjunction with weaknesses. Having both strengths and weaknesses in diagnostic feedback can promote learning among students who receive feedback and encourage growth in their learning trajectories. As for teachers, this all-round diagnosis can help them both in providing feedback and designing remedial activities for students.

Lee (2015) conceptualizes three major components in designing and developing DLA tools and procedures, including: 1) *diagnosis*, 2) *feedback*, and 3) *remedial learning and instruction*. Seen as the core component of DLA, diagnosis involves identifying the strengths and weaknesses that prevent a student or a group of students from progressing to the next level or stage of development. In addition, the diagnosis component does not only focus on listing strengths and weaknesses, but also on finding the causes of weaknesses so that further action can be taken appropriately. The second element of DLA is feedback. This refers to summary diagnoses for both learners and teachers. The presentation of diagnostic feedback is crucial, as strengths and weaknesses should be communicated to both learners and teachers in an effective and comprehensible manner so that they can follow the recommendations made. Importantly, for DLA to be learning-oriented,

feedback should allow students not only to understand their current level, but also to *feedforward* into their future learning.

The last component is remedial learning and instruction. This means that once students' weaknesses are identified, they should be provided with learning activities or advisory sessions that can help them address their weaknesses and achieve their learning goals. Lee (2015) outlines three types of remedial learning: remedial activities provided by DLA developers, remedial programs designed by teachers and/or institutions, and self-learning materials that students can use independently. The decision to choose certain remedial activities can vary depending on the context and course/curriculum objectives. For example, in contexts where students are generally self-directed and the content of the course is more open-ended, self-help remedial activities might be designed and implemented. However, if DLA is designed, developed, and implemented in close connection with a school curriculum and syllabus, a teacher-led remedial program might be more suitable, as teachers can collaborate with administrators/institutions to offer support to students. Despite these components being theoretically sound, there remain some challenges that teachers wanting to implement DLA locally might have to face. First, there seems to be no clear specification as to how DLA instruments and procedures should be constructed, how feedback should be provided, and how these can be linked to remedial learning. The second challenge relates to ensuring that diagnostic feedback is sufficiently effective and learning-oriented for students to feedforward into their learning. It is still unclear how feedback should be constructed and what type of feedback would be most facilitative for learning.

These challenges are valid and should be addressed by the language assessment community. However, when it comes to the actual use of DLA in the language classroom, the challenges are more about practical issues. Many teachers would like to provide their students with detailed diagnostic feedback, but everyone is well aware how laborious that is. With ChatGPT, the process can

be partially automated so that teachers can focus more on delivering feedback and designing remedial help for students. Theoretical developments and debates aside, I took the liberty of adopting a syllabus-based approach to DLA (Bachman, 1990) and designed a ChatGPT-based self-diagnostic assessment based on the objectives of the course I was teaching. This syllabus-based approach focuses on aligning assessments closely with the content and objectives outlined in a course syllabus. It was chosen not only because it was the most appropriate approach in my context, but also because it was the most practical.

3. Using ChatGPT for Diagnostic Writing Assessment

I adopted ChatGPT as an assistant tool for conducting diagnostic writing assessments for English-major students enrolled in a paragraph writing course at a public university in Thailand. All 30 students in class had reasonably good levels of English proficiency, ranging from B1 to B2 based on the in-house English proficiency test employed by the university. It should be noted that writing was tested through error identification and multiple-choice questions, but not writing production. Hence, their writing ability still remained to be seen. In the first class, I asked my students to have their devices with them (i.e., laptops or iPads). There were four main steps involved in utilizing ChatGPT for diagnostic writing assessment.

3.1 Step 1: Creating Instruments

Two types of prompts were created: writing prompts and ChatGPT prompts. The first step was to create a diagnostic instrument, a writing prompt that elicited answers in a paragraph format. The prompt question needed to reflect the types of writing that students were expected to learn in class (e.g., descriptive paragraphs, process paragraphs, etc.). In my case, I used the question “What is your day to remember?”. The next step was to identify elements of writing components (subskills) that should be diagnosed based on the actual rubrics used to assess students in class. The rubrics included grammar (sentence structure and

accuracy), vocabulary, organization, mechanics, and task response. Other prompts that students needed to use, such as asking ChatGPT to identify strengths and weaknesses, were also written. These ChatGPT prompts were prepared in a ready-to-use format for students to use as they moved from one step to another. A separate Google Document was also prepared for students to input their diagnostic results and profiles. A total of five prompts were created: one writing prompt and four ChatGPT prompts. They are shown in Table 1 below:

Table 1

Prompts Used

Writing Prompt	ChatGPT prompts
What is your day to remember?	<p>Can you provide a diagnostic writing assessment for a written response of the question “What is your day to remember?” based on sentence structure, grammatical accuracy, vocabulary, organization, mechanics, and task response?</p> <p>Can you provide detailed feedback on each of the criteria?</p> <p>Can you create a table showing the strengths and weaknesses in this piece of writing?</p> <p>Can you create a learner’s writing skills profile in a table format based on the feedback, as well as the strengths and weaknesses identified above?</p>

3.2 Step 2: Diagnosis

There were two phases in this step: a preparation phase and a diagnostic phase. In the preparation phase, the students were asked to create a ChatGPT account. They were then provided with prompts and instructions. In the diagnostic phase, they were asked to complete the tasks within 20 minutes. After they finished writing, they had to use the provided prompts to ask ChatGPT for a diagnosis (see Figure 1).

Figure 1

A ChatGPT Prompt for Diagnosis

"Can you provide a diagnostic writing assessment for a written response of the question "What is your day to remember? based on sentence structure, grammatical accuracy, vocabulary, organisation, mechanics and task response.

The first day of entering my dorm wasn't fun at all. It was a very busy morning and it took a long time to travel. I had milk and bread for breakfast. I got car sickness. It was really bad. I vomited in the car until it stained my clothes. Then my shirt smelled of spoiled milk and stinks all day. When I reached the dorm, I went up to the dorm but walked to the wrong building. It was a very dizzy morning. In the evening, there are activities with seniors to find a family similar to the code line. It's a day with a variety of moods, there are things to be excited about all day, but it's very fun.



3.3 Step 3: Feedback

After receiving this initial diagnosis, the students used the provided prompts to elicit detailed feedback on each criterion (Can you provide detailed feedback on each of the criteria?) and asked ChatGPT to summarize the feedback in a table format (see Figures 2 and 3).

Figure 2

Detailed Feedback on Sentence Structure



ChatGPT

Sure, I'd be happy to provide detailed feedback on each of the criteria for the written response.

1. Sentence Structure:

The response contains some errors in sentence structure. The sentences tend to be short and simple, and lack variety in sentence structure. The use of subordinating clauses could have helped to vary the sentence structure and make the writing more engaging. The writer could also have used more complex sentence structures, such as compound-complex sentences, to convey their ideas more effectively.

Example of improvement:

Instead of writing: "It was a very busy morning and it took a long time to travel," the writer could write: "Despite the fact that it was a very busy morning, it took a long time to travel to my dorm."

Figure 3*Summary of Feedback Based on Each Criterion***ChatGPT**

Sure, here's a summary table of the feedback for each criterion:

Criteria	Feedback
Sentence Structure	Short and simple sentences; lack of variety in sentence structure; use of subordinating clauses and complex sentence structures could improve writing
Grammatical Accuracy	Errors in tense usage; lack of attention to subject-verb agreement, singular/plural forms, and article usage
Vocabulary	Simple and repetitive vocabulary; more descriptive and engaging vocabulary could improve writing
Organization	Lack of coherence and organization; use of transitional words or phrases could improve flow of writing
Mechanics	Errors in punctuation and inconsistent capitalization; attention to spelling and formatting could improve readability
Task Response	Adequate response to prompt but could use more detailed and engaging information, personal insights, and reflections.



As it is important for a DLA to focus not only on weaknesses but also on strengths, the students then asked ChatGPT to identify the strengths and weaknesses of their writing using the provided prompt (Can you create a table showing the strengths and weaknesses in this piece of writing?), as shown in Figure 4.

Figure 4*Summary of Strengths and Weaknesses***ChatGPT**

Sure, here's a table showing the strengths and weaknesses of the written response:

Strengths	Weaknesses
Provides a clear response to the prompt	Sentence structure lacks variety and complexity
Contains some descriptive language	Grammatical errors and inconsistencies
Conveys a range of emotions throughout the day	Vocabulary is simple and repetitive
Includes personal experience and reflection	Lack of coherence and organization
Includes some details and specific examples	Errors in punctuation and capitalization
	Could provide more engaging and detailed information
	Could use more precise and specific vocabulary
	Could use transitional words or phrases to improve flow
	Could pay closer attention to spelling and formatting

The next step was for the students to ask ChatGPT to create their own writing skills profiles based on the feedback provided (Prompt: Can you create a learner's writing skills profile in a table format based on the feedback and strengths and weaknesses identified above?), as shown in Figure 5.

Figure 5*A Writing Profile*

Writing Skill	Strengths	Weaknesses	Improvements
Sentence Structure	Uses short and simple sentences	Lacks variety and complexity in sentence structure; could benefit from using subordinating clauses and more complex sentence structures	Practice incorporating subordinating clauses and complex sentence structures; vary sentence structure for more engaging and varied writing
Grammatical Accuracy	Conveys a clear message and provides some descriptive language	Errors in tense usage; lack of attention to subject-verb agreement, singular/plural forms, and article usage	Review grammar rules and practice identifying and correcting errors in writing
Vocabulary	Provides some descriptive language and conveys a range of emotions	Vocabulary is simple and repetitive; could use more precise and specific vocabulary	Expand vocabulary by reading more widely and looking up new words; use more precise and specific vocabulary to enhance writing
Organization	Includes personal experience and reflection	Lack of coherence and organization; could use transitional words or phrases to improve flow	Use transitional words or phrases to create a clear and logical flow in writing; consider outlining or pre-writing to organize ideas before starting to write
Mechanics	Conveys a range of emotions throughout the day	Errors in punctuation and inconsistent capitalization; could pay closer attention to spelling and formatting	Review punctuation and capitalization rules; proofread carefully to catch spelling and formatting errors
Task Response	Includes specific examples and personal reflection	Could provide more engaging and detailed information; could use more descriptive language ↓	Provide more detailed and engaging information by including specific details and sensory language; use personal reflection to deepen insights and reflections

Finally, the students were asked to copy the conversation thread with ChatGPT and paste it into a Google document previously prepared for them.

3.4 Step 4: Using Reports for Remedial Learning

The students' profiles and feedback were reviewed to determine if they accurately reflected the students' writing. I then summarized common areas of improvement, as well as identified students who might need more support so that a plan for remedial learning could be devised. An improvement summary was presented to the students in class as group feedback. The students were also asked to use the feedback to help with their subsequent writing in class.

4. Reflections and Implications

To me, ChatGPT was very effective in providing diagnostic feedback for my students. When cross-checking the feedback with their actual writing, the program was able to spot relevant problems the students had in their writing fairly accurately. In my personal assessment, the feedback generated by ChatGPT was generally accurate in all criteria and for all students. Moreover, the program was able to summarize strengths, weaknesses, and areas of improvement in a ready-to-use format (table), something that not only saved time for teachers, but also was more accessible for students compared to traditional paper-annotated feedback. Some of my students mentioned that this approach was conducive to independent learning, and that they adopted the same approach for other subjects. There are, however, some caveats for language teachers considering using ChatGPT in the same way. First, my students were English majors with a good level of English; therefore, they were able to make sense of the summarized feedback with relative ease. Also, as language students, they were able to understand grammatical terms used in the feedback fairly well. Students of lower proficiency level (A1-A2) might need more assistance in understanding and making sense of their profiles. Prompts specifying the use of non-technical terms can also be used. Of course, you can sit your students down and explain the feedback to them one by one, but that would require a great amount of labor and time, which some teachers may not have in abundance. What I recommend, then, is to add another prompt at the end of Step 3 (feedback), asking ChatGPT to translate the

summarized feedback into the students' first language so that they can fully understand what they need to improve.

Another caveat is the subject matter you are teaching. Using ChatGPT to provide detailed diagnostics at the paragraph level appears effective, but I am not certain that it would extend to longer texts (e.g., essays or reports) or other types of writing (e.g., creative writing). Nevertheless, given the large amounts of data that ChatGPT draws on, I speculate that it would be equally effective. The final concern is an ethical one. Now that you have let the genie out of the bottle, it is likely that the students will use the same approach in subsequent graded writing assignments. Hence, a common ground regarding how and to what extent ChatGPT and other writing-assistant tools can be used to facilitate the writing process must be thoroughly discussed. My final thought on using ChatGPT as a language diagnostic tool, and the hype around it in general, is that students with a high level of language proficiency (or proficiency in a given subject) have a better tendency to capitalize on the benefits of the program than those with lower proficiency. This is because they need to be able to not only understand what they read, but also evaluate whether the feedback actually makes sense. Their knowledge of what they seek help from ChatGPT must be superior to, or at least on a par with, it. Otherwise, students may run the risk of seeking unethical help from AI and learning nothing at the end of the day.

5. About the Author

Napat Jitpaisarnwattana (www.digital-elt.org) is a lecturer of English at the Department of Western Languages, Faculty of Archaeology, Silpakorn University. His research interests are Language MOOCs, AI in language education, learning analytics, digital wellbeing, and computer-assisted language learning (CALL). He can be reached at napat.jit@gmail.com

6. References

- Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Toward a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, *36*(2), 236–260.
<https://doi.org/10.1093/applin/amt046>
- Al-Mukdad, S. (2019). Investigating English academic writing problems encountered by Arab International University students. *Theory and Practice in Language Studies*, *9*(3), 300–306.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 1–20.
<https://doi.org/10.1007/s10639-023-11834-1>
- Jitpaisarnwattana, N. & Saville, N. (2024). Technology and learning-oriented assessment. In L. Liantas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2)
<https://doi.org/10.1177/00336882231162868>
- Kunnan, A. J., & Jang, E. E. (2011). Diagnosing feedback in language assessment. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 610–627). Wiley-Blackwell.
- Lee, Y.-W. (2015). Diagnosing diagnostic language assessment. *Language Testing*, *32*(3), 299–316.
- Nurbayan, Y., Al Farisi, M. Z., Sanusi, A., & Supriadi, R. (2020). Investigating the students' weaknesses in writing academic papers. In *4th International Conference on Language, Literature, Culture, and Education (ICOLLITE 2020)* (pp. 785–791). Atlantis Press.
<https://doi.org/10.2991/assehr.k.201215.124>

- Nguyen Minh, A. (2024). Leveraging ChatGPT for Enhancing English Writing Skills and Critical Thinking in University Freshmen. *Journal of Knowledge Learning and Science Technology*, 3(2), 51–62.
<https://doi.org/10.60087/jklst.vol3.n2.p62>
- Özçelik, N., & Ekşi, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—a case study. *Smart Learning Environment*. 11(10).
<https://doi.org/10.1186/s40561-024-00296-8>
- Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). ‘To generate or stop generating response’: Exploring EFL teachers’ perspectives on ChatGPT in English language teaching in Thailand. *Learning: Research and Practice*, 9(2), 168–182. <https://doi.org/10.1080/23735082.2023.2257252>
- Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). ChatGPT in language education: How to use it ethically? *PASAA*, 67, 361–377.