International Journal of Chinese Language Teaching (2024)
Vol. 5 (1) 74-89 https://doi.org/10.46451/ijclt.20240106

Article

Utilizing ChatGPT to Implement Differentiated Instruction

Qiuchen Li
Jiafan Zhang
Wei Cai*
University of Calgary, Canada

Received: 10 August, 2023/Accepted: 17 September, 2023/Published: 20 February, 2024

Abstract
The study explores the potential of using ChatGPT in facilitating differentiated instruction, focusing on its ability to assess Chinese learners' language abilities, produce materials in different genres and at different levels, create teaching tasks, and simulate assessments. The correlation was calculated between the original scores and ChatGPT-generated scores of forty-five randomly selected HSK test writing samples. Additionally, ChatGPT's ability to generate diverse materials was tested by simulating thirty texts across various genres and levels. The study also examined ChatGPT's capability in creating a range of tasks and assessments.

The result showed a significant correlation between the original scores and those generated by ChatGPT, indicating its ability as a useful tool to measure learners' language performance. ChatGPT demonstrated efficacy in generating materials spanning different genres and difficulty levels, aligned with the CEFR benchmarks. Given specific and well-structured prompts, ChatGPT proved adept in tailoring tasks and assessments. Further research is crucial to understand the application of ChatGPT in differentiated instruction.

Keywords
ChatGPT, differentiated instruction, Chinese learners, language abilities

1 Introduction

ChatGPT has sparked significant interest in the field of language learning given its potential to enhance both students' learning outcomes and the efficacy of language instruction (Kostka & Toncelli, 2023). Current discussions on ChatGPT in second language (L2) teaching focuses on language skills (Barrot, 2023; Yan, 2023), effects on language teaching or learning (Hong, 2023; Rudolph et al., 2023), and analyses of the strengths and weaknesses of ChatGPT (Kohnke, 2023; Kuhail et al., 2023). We should further explore and experiment with the use of ChatGPT to facilitate varied approaches in language teaching, one of which is differenced instruction. Traditional teaching models often adopt a one-size-

*Corresponding author. Email: wcai@ucalgary.ca

fits-all approach, overlooking students' diverse backgrounds, learning styles, and proficiency levels. The incongruence between teaching methodologies and student diversity has spurred the rise of differentiated instruction – a pedagogical approach tailored to learners' differences (Tomlinson, 2003). ChatGPT has a range of capabilities: generating texts, developing lesson plans and assessments, providing immediate feedback, fostering learner interaction, and so on (Cai, 2023 a, b; Kohnke et al., 2023; Kasneci et al., 2023). These capabilities position ChatGPT as a promising tool to facilitate differentiated instruction.

The essence of differentiated instruction lies in its ability to empower educators to tailor their teaching strategies and materials to suit the individual learning characteristics of their students (Tomlinson & Eidson, 2003). By recognizing the unique attributes of each learner and adapting instructional content, processes, and assessments accordingly, differentiated instruction endeavors to unlock the utmost potential of individual students, irrespective of their backgrounds or abilities. For instance, some students might thrive in visual learning environments (such as videos, photos), while others excel through hands-on activities or discussions. Through differentiation, educators can provide varied modalities of instruction that resonate with each student's learning preferences. This approach not only enhances student engagement and motivation but also promotes holistic development by addressing cognitive, affective, and skill-based dimensions of learning (Fisher & Frey, 2008; Gregory & Chapman, 2013).

Although the concept of differentiated instruction is widely recognized, teachers may find it difficult to implement it due to lack of time and insufficient resources (Gibbs, 2023). This article delves into the applicability of ChatGPT to facilitate differentiated instruction. Through a comprehensive exploration of four key aspects – assessment, content and materials, process, and product – this research aims to shed light on how ChatGPT can contribute to enhancing differentiated teaching. By examining both the strengths and shortcomings of integrating ChatGPT into differentiated instruction, this study seeks to provide educators with balanced perspectives on leveraging ChatGPT to optimize their pedagogical practices.

2 Differentiated instruction in language teaching

Differentiated instruction focuses on students' needs, goals, and individual differences, requiring teachers to design and adjust the teaching content, process and product based on students' proficiency levels, interests and learning styles (Bondie & Zusho, 2018). Such an approach proves especially beneficial for Chinese language classes in North America, given the diverse backgrounds of the students. Unless adopting a two-track curriculum, Chinese classes in North America typically accommodate both heritage and non-heritage language learners who differ in terms of language proficiency, affective needs, learning goals, and learning methods. An effective educator acknowledges the potential impact of these disparities on the learning process and strives to establish a classroom environment that appreciates and respects this diversity (Tileston, 2004). The diversity of students in the classroom requires differentiated instruction (Shi & Hua, 2007). It plays a crucial role in addressing the unique learning needs of each individual student (Bi et al., 2023).

In a shared classroom environment, differentiated instruction organizes instructional materials, activities, and assessments in a way that students with different characteristics can achieve their individual goals and attain holistic development. It transforms traditional teaching models into dynamic, enjoyable, autonomous, and effective learning experiences (Birnie, 2015; Tomlinson, 2005). As a learner-centered approach, differentiated instruction allows educators to customize their teaching methods to individual learners, thereby fostering greater student engagement and motivation (Moosa & Shareefa, 2019; Tomlinson & Allan, 2000). Researchers identify differentiation of content, process, and product as the three crucial dimensions of differentiated instruction (Blaz, 2016; Chien, 2012).

2.1 Differentiated teaching content and materials

One way of differentiating instruction is to differentiate teaching content and materials based on the assessment of students' proficiency levels, background knowledge, interests, and other relevant factors. This underscores the importance of an initial assessment of learners' backgrounds. ChatGPT could potentially enhance the efficiency of these assessments, such as students' language levels. Additionally, instructors can provide students with open-ended learning content. In language teaching, instructional content encompasses not only linguistic aspects such as vocabulary and grammar but also non-linguistic dimensions like attitudes, skills, and strategies (Blaz, 2006).

By implementing differentiated content instruction, students can engage with the provided materials at a pace tailored to their specific satiations. Once students have fulfilled their obligatory assignments, they gain the autonomy to opt for other materials or tasks aligned with their personal interest. Consequently, every student proceeds to engage in distinct materials and associated tasks, whether independently, in collaboration with peers, in group settings, or with teacher guidance (Chien, 2012).

Blaz (2016) suggests three ways to help students engage with different instructional materials. The first entails differentiation based on students' interests, allowing each individual to explore their own areas of curiosity. Subsequently, in smaller groups or through online platforms, students can engage in the sharing and comparison of their discoveries. The second way involves furnishing uniform materials to all students, offering different reading levels and different degrees of words. The third method to differentiate delivery of content involves giving students choices in the type of instruction: direct instruction, worksheet practice, online work or more complex activities (Blaz, 2016). This third method essentially involves differentiation of the teaching process, which we will explore in the next section.

While differentiated instructional materials are beneficial to students' academic progress, the challenge lies in selecting and gathering the appropriate content and materials for different students. This process can be time-consuming. Recent studies have shown that ChatGPT can generate customized learning resources that cater to different learning preferences and levels (Kasneci et al., 2023). It could be instrumental in developing differentiated materials when provided with specific prompts.

2.2 Differentiated teaching process

In differentiated instruction, the teaching process refers to the process through which students comprehend information, ideas, and skills, typically through various activities or tasks (Blaz, 2006). Subsequently, Blaz (2016) proposed that the term "process" should also include the various ways in which students make sense of the content or input. For differentiated instruction to be effective, teachers should factor in students' interests, cognitive abilities and learning styles (Chien, 2012). As such, the process should cater to the diverse learning styles of students (Blaz, 2016). Given the differences among students, the teaching process should also be flexible and diverse, often requiring the integration of multiple means and methods within a single class. When implementing differentiated teaching processes, teachers initially need to understand students' interests, cognitive abilities, and learning styles. Based on this understanding, they can conduct various classroom activities, introduce different learning strategies, and select impactful methods to convey concepts, information, ideas, and skills to students (Chien, 2012) For instance, teachers can flexibly choose between whole-class instruction, group instruction, or individualized instruction based on students' situations, allowing for homogeneous grouping or heterogeneous cooperation (Shi & Hua, 2007).

Despite the potential benefits of differentiating the process, instructors often dedicate a significant amount of time to planning diverse assignments, activities, task formats, learning objectives, assessment methods, and other related aspects. Given its powerful capabilities to generate texts, activities or tasks, it is worth exploring how we can utilize ChatGPT to differentiate processes.

2.3 Differentiated product

Differentiating the product involves varying the complexity of what students produce to demonstrate their level of mastery of the unit's content. Products may be formal (e.g., a report) or informal (e.g., an interview) (Blaz, 2016). Lower-level students can tackle relatively simpler tasks, while higher-level students engage in more complex ones. However, for each individual student, the assigned tasks should be challenging, critical, and innovative in nature. Therefore, to cater to the diverse needs and abilities of students, tasks should be adapted to their individual proficiency levels, learning styles, and preferences.

Assessments should encompass a holistic evaluation of students, extending beyond traditional exam scores (Shi & Hua, 2007). L2 instructors must differentiate between difficulty levels in assessments, employ diverse evaluations, and offer multiple avenues for students to demonstrate learning achievements. For example, formative assessments pay attention to the process of the evaluation, whereas summative assessments emphasize the product (Nieminen & Tuohilampi, 2020).

In conclusion, differentiated instruction offers the advantage of enhancing students' learning outcomes and can be implemented through the differentiation of content, processes, and product. Differentiated assessment approaches, which are aligned with student's skill levels and learning profiles, provide opportunities for varied demonstrations of knowledge and competencies. A fundamental step for differentiated instruction is the pre-assessment of learners' language abilities and other characteristics. To examine the potential in using ChatGPT for differentiated instruction, the current study aims to address the following research questions:

1. Is there a correlation between the original HSK writing scores and those generated by ChatGPT?

   This research question explores the feasibility of employing ChatGPT as a tool for the preliminary assessment of learners' language skills, specifically their writing skills.
2. Is ChatGPT capable of generating materials that progress in difficulty and vary in genres, aligned with CEFR benchmarks?
3. Can ChatGPT create tasks in line with Ellis's (2003) framework?

   This research question examines ChatGPT's capability for facilitating process differentiation.
4. Is ChatGPT able to simulate different types of assessments?

   This research question probes into ChatGPT's potential for contributing to product differentiation.

By addressing these questions, the study aims at evaluating ChatGPT's capability to assess Chinese learner's writing performance, to generate content and materials of different genres and difficulty levels, to create tasks of various types, and to simulate assessments.

3 Procedure

Firstly, students' written products may be graded by ChatGPT, provided its credibility as a reliable grading tool is verified. To show its reliability, this study randomly selected 45 written samples from the HSK Dynamic Composition Corpus 2.0, which collected some Chinese learners' written products of the HSK test from 1992 to 2005. These samples were chosen for their diverse range of nationalities, writing topics, and grades. Fifteen written products were chosen from each level of the HSK test: A, B and C. Moreover, writings from four time slots in 2005, the most recent examination year in the database, were included. Before deploying ChatGPT 4.0 for grading, the grading rubric, which was based on the writing task of the HSK test, and a grading range from 0 to 100 was provided for ChatGPT. The prompt "Please

grade the writing based on the rubric" was introduced whenever the topic of the written sample changed to ensure that ChatGPT adhered to the rubric.

Secondly, ChatGPT presents a potential solution for generating content or materials of various genres and difficulties, but it is crucial to examine the genre and text complexity of generated content or materials before deployment. The present study used the genres and difficulty levels of the Common European Framework of Reference (CEFR) to evaluate ChatGPT's ability of generating appropriate genres and difficulty levels. CEFR categorizes genres or text types into description, narrative, instruction, explanation, and argumentation, which aligns with the Common Core State Standards in the US national curriculum (Natova, 2021). Furthermore, CEFR identifies six difficulty levels: A1, A2, B1, B2, C1, and C2. Using this framework, we tasked ChatGPT with creating thirty Chinese texts on the topic "climate change", spanning all the above genres and difficulty levels (e.g., description A1, A2, B1, B2, C1, and C2). The description of each difficulty level, based on CEFR, was inputted first to clarify the requirement of the difficulty. All texts were within a 500-character limit. Analyses were performed using L2C Rater and Common Text Analysis Platform (CTAP), tools for evaluating text complexity.

Thirdly, using Ellis's (2003) task framework, we evaluated ChatGPT's ability to distinguish different types of tasks. Ellis (2003) categorizes tasks into focused and unfocused tasks, with the former honing learners' ability to understand linguistic features and the latter emphasizing comprehension and production of the language for communication. Focused tasks are subdivided into comprehension, structure-based production, and consciousness-raising tasks, targeting the comprehension of grammatical structures, the use of grammatical forms in a natural context, and fostering grammar awareness. To examine ChatGPT's capabilities in distinguishing the tasks, prompts were created, specifying the topic and type of the task as well as the proficiency of target learners. The prompts were experimented and refined to examine the influence of prompts on the task quality. Moreover, prompts were also provided to check ChatGPT's ability to generate vocabulary lists and to choose a mandatory or optional task from a list of tasks so that teachers can assign the mandatory task for learners to complete or ask learners to choose the task that they are interested in.

Fourthly, ChatGPT's capability to provide assessments is examined, as offering diverse types of assessments helps address learners' learning styles, interests, strengths and weaknesses, meeting the basic requirement of differentiated instruction (Blaz, 2016). If ChatGPT can identify the characteristics of different types of assessments and generate them accordingly, teachers may use these assessments as a foundation, requiring only minor revisions. This could considerably save time in creating various types of assessments. Different prompts were inputted, and trials were made to examine ChatGPT's competence in generating formative and summative assessments[1]. The capacity of ChatGPT to design form-focused and meaning-focused assessments, aligning with FonFs (focus on forms) and FonF (focus on form)[2], was examined. Including these two types of assessments is consistent with the important roles they each play in L2 learning. The prompts provided to ChatGPT specified the topic of climate change, and the targeted learners across the beginning, intermediate and advanced levels.

To answer the four research questions, the study used SPSS and Excel to examine the capability of ChatGPT in gauging learners' writing performances, generating teaching materials, creating teaching tasks and assessments.

4 Results

Results of the study are reported corresponding to each of the research questions.

Research question 1: Is there a correlation between the original HSK writing scores and those generated by ChatGPT?

The scores given by the original raters of each written product were compared with those generated by ChatGPT. A Pearson correlation was calculated to determine the relationship between the two sets of scores. The result shows a fair degree of relationship between the ratings (r = .41). See Table 1.

Table 1

Correlations Between Original Score and Score of ChatGPT

		Original score	Score of ChatGPT
Original score	Pearson Correlation	1	.405**
	Sig. (2-tailed)		.006
	N	45	45
Score of ChatGPT	Pearson Correlation	.405**	1
	Sig. (2-tailed)	.006	
	N	45	45

**. Correlation is significant at the 0.01 level (2-tailed).

Research question 2: Is ChatGPT capable of generating materials that progress in difficulty and vary in genres, aligned with CEFR benchmarks?

To answer this research question, L2C Rater assessed the difficulty levels of the thirty texts across six genres, providing an overall score based on the evaluation of content, lexical richness, lexical complexity, variety, and complexity of phrase structures. As illustrated in Figures 1 to 5, as the difficulty levels increased, scores generally increased across the five genres; there are outliers, especially at advanced levels including B2, C1 and C2. It should be mentioned that in some genres, as presented in Figure 1, texts of C2 level, even though equating to C1 in the overall scores, surpassed texts of C1 in lexical richness.

Figure 1

Difficulty Level and Score by L2C Rater of Description

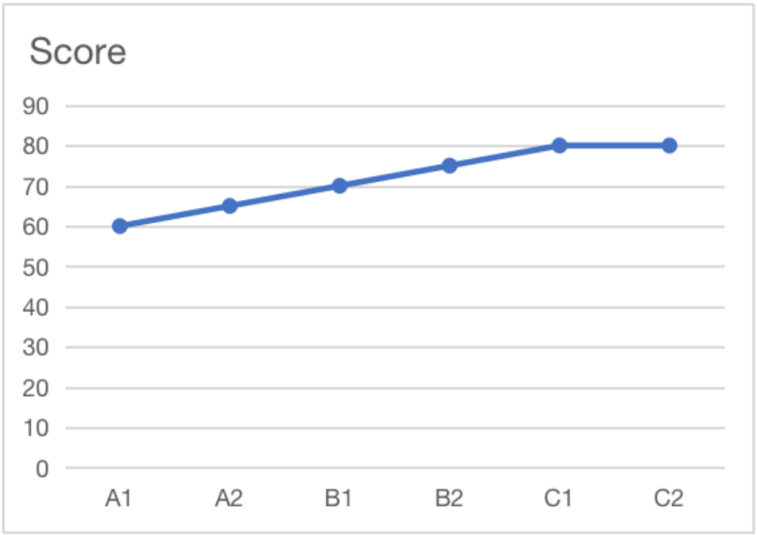


Figure 2

Difficulty Level and Score by L2C Rater of Narrative

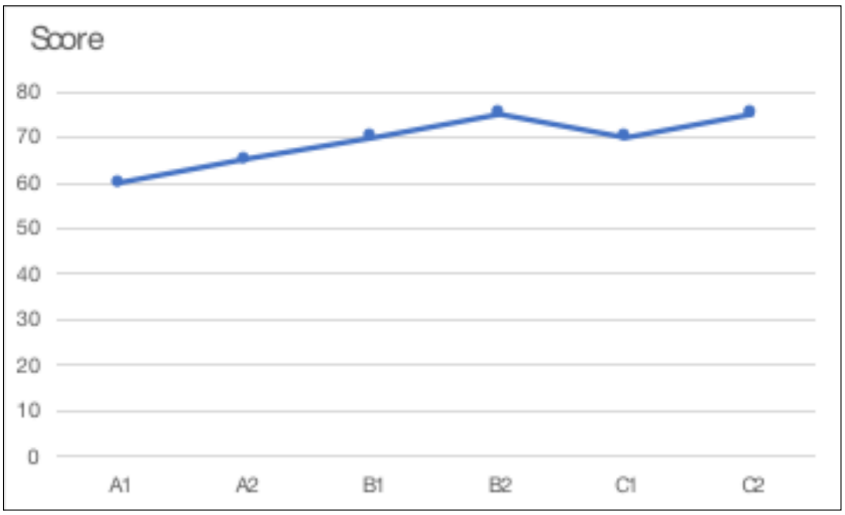


Figure 3

Difficulty Level and Score by L2C Rater of Instruction

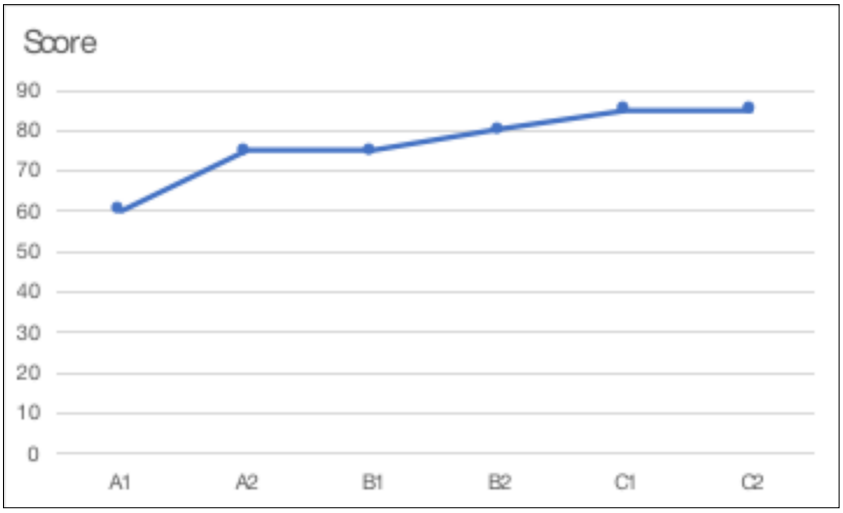


Figure 4

Difficulty Level and Score by L2C Rater of Argumentation

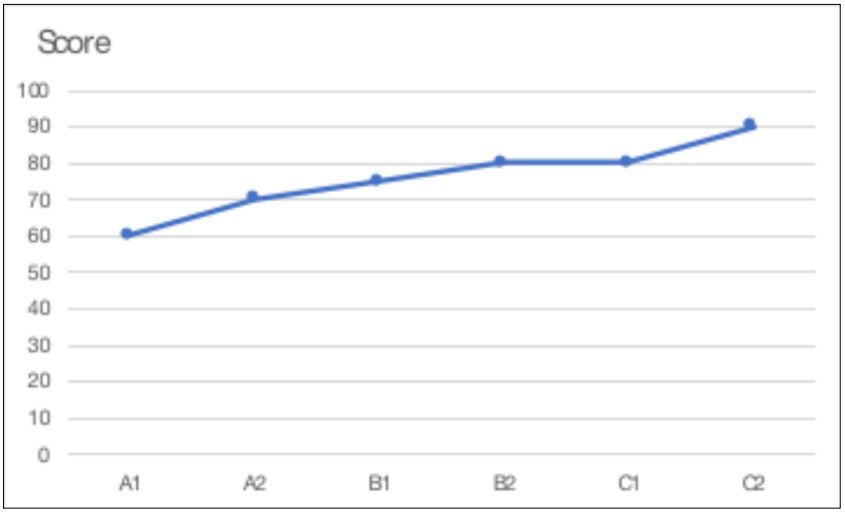
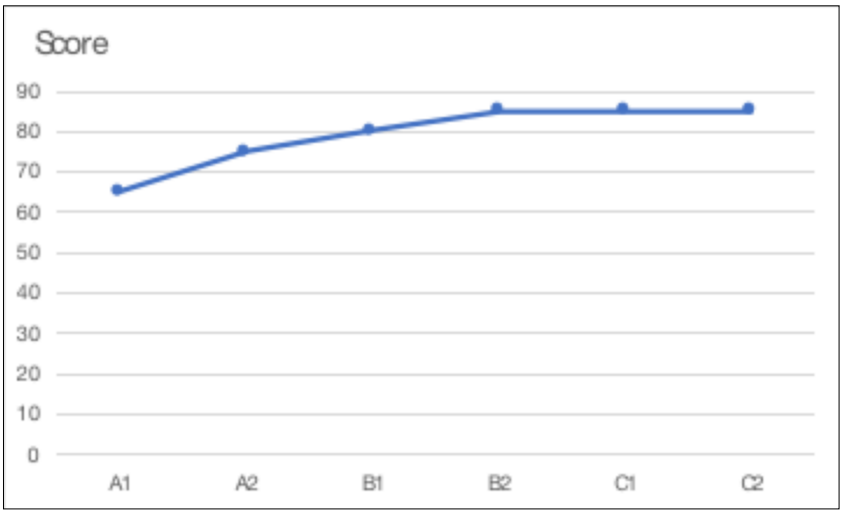


Figure 5

Difficulty Level and Score by L2C Rater of Explanation



A detailed analysis of the texts was conducted by CTAP to evaluate the text complexity from three aspects, encompassing 12 features: character richness includes the number of tokens, number of types, and type token ratio; vocabulary includes the average word length, lexical richness (measured by the type token ratio), lexical variation feature (noun), and the number of two-character words; sentence includes the average sentence length based on characters, the length of the longest sentence (in both character and word length), and syntactic complexity (indicated by the mean length of noun phrases and the number of noun phrases per sentence). The result shows that the indices of the 12 features, which capture the text complexity from the aforementioned three aspects, generally increase in tandem with the increasing difficulty levels across all genres. For example, Figures 6, 7 and 8 display character richness in terms of tokens, types, and type token ratios, revealing the match between the measured complexity of texts and the difficulty levels from A1 to C2. Similar results are also observed in other features pertaining to vocabulary, including the average word length, lexical richness (the type token ratio), lexical variation of nouns, and the number of two-character words. Similar trends are evident for sentences, such as noun phrases per sentence and the mean length of noun phrases.

Figure 6

Character richness: Number of Tokens

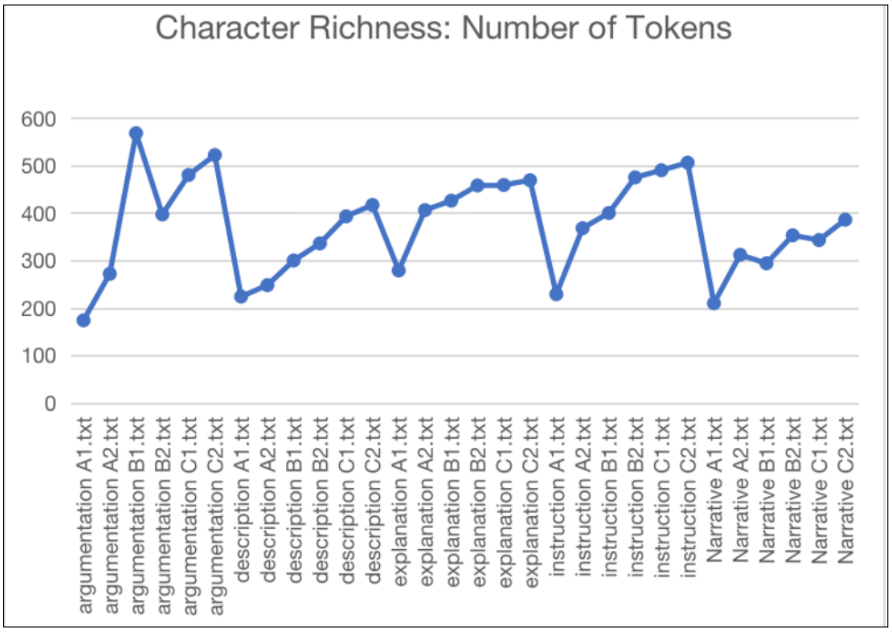


Figure 7

Character Richness: Number of Types

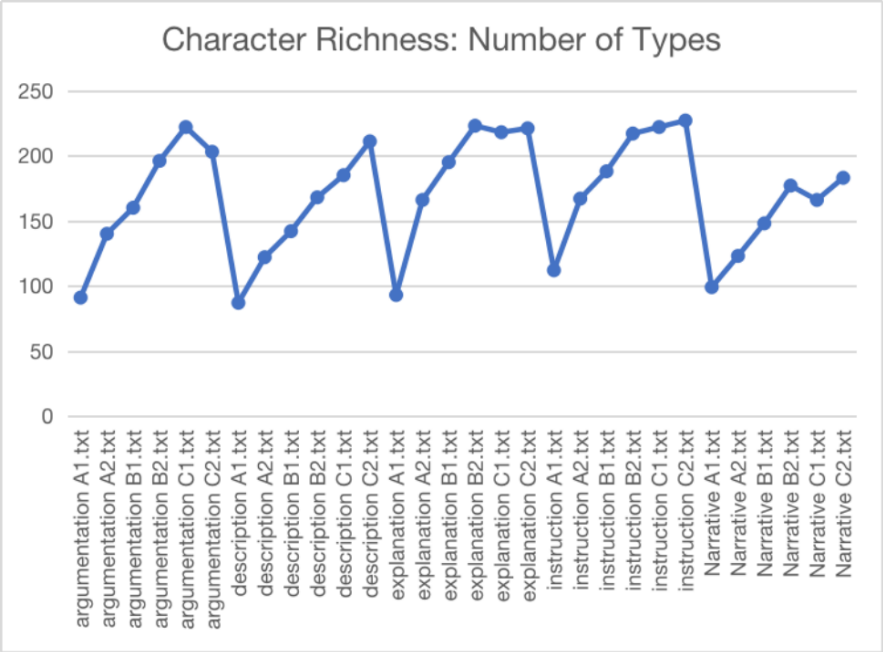
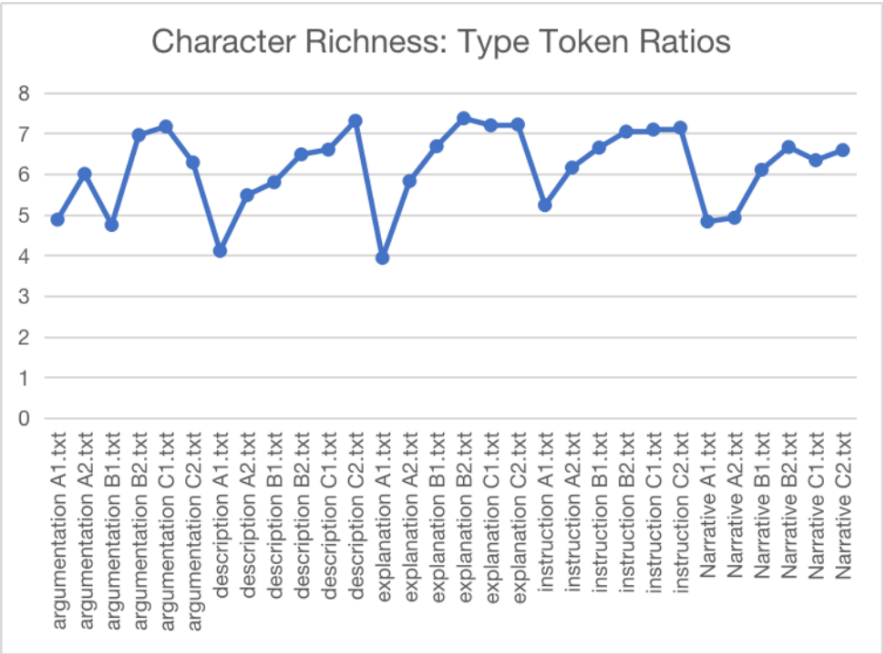


Figure 8

Character Richness: Type Token Ratios



Research question 3: Can ChatGPT create tasks in line with Ellis's (2003) framework?

Drawing from Ellis' framework, this research question examines ChatGPT's capabilities in generating specified tasks. The result shows that ChatGPT adeptly creates various types of tasks, aligning with Ellis's (2003) categorizations and matching the described characteristics, as presented in Table 2. For instance, an unfocused task was generated to enhance learners' understanding of climate change and global environmental issues, which aims at improving learners' ability to comprehend the language of the topic for communicative purposes.

Table 2
Features of Tasks Generated by ChatGPT

Tasks	Definitions or descriptions	Features	Features of tasks produced by ChatGPT
Unfocused tasks	focusing on comprehension and production of language for purposes of communication	comprehension or production of the language in a certain topic	√
		communication as the major purpose	√
		not eliciting attention to specific linguistic features	√
Comprehension tasks	helping learners attend to target grammar forms contained in the given input	focusing on target grammar forms	√
		comprehending not producing the structure	×*
		task containing a stimulus to respond nonverbally or with minimal use of the target language	×*
Structure-based production tasks	enabling learners to perform a communicative activity by using a target grammar form naturally, usefully or essentially	using the target grammar to conduct the communicative activity	√
		utilizing the target grammar is natural, useful or essential.	√
Consciousness-raising tasks	providing learners with input containing examples of target grammar forms to operate on the input	input embracing examples of target grammar forms	√
		operating on the input to develop awareness of grammatical properties	√

Note: Features are based on Ellis (2003).

It should be noted that the comprehension task lacked two features in the trials when the prompts did not describe them. However, upon refining and including these features in the prompts, the generated tasks can incorporate these features.

ChatGPT can generate a variety of task formats, including debates, storytelling, role-playing, presentation, and group discussions. For each task, ChatGPT provides a detailed instruction, outlining the task title, objectives, procedures, evaluation criteria, and resources. The instructions are clear and can guide students systematically through each step, from research and preparation to writing a position statement, preparing for the debate and concluding classroom discussion.

Research question 4: Is ChatGPT able to simulate different types of assessments?

The trials show that with a detailed prompt specifying the assessment type, topic, and learners' proficiency, ChatGPT can generate different forms of assessments accordingly, such as quizzes and presentations (see Table 3). When provided merely with the name of the type, either summative or formative, ChatGPT can recognize characteristics of each and generate the desired assessment.

The result also suggests that ChatGPT is able to generate form-focused and meaning-focused assessments appropriate for the given topic and intended learners.

Table 3

Features of Assessments Generated by ChatGPT

Assessments	Definitions or descriptions	Features	Features of assessments produced by ChatGPT
Summative assessments	assessments that focus on the end result or the achievement mostly upon the completion of the course	to evaluate students' achievements or learning outcomes	√
		usually at the end of the course	√
		to evaluate grades toward benchmarks	√
Formative assessments	assessments that use the results mainly to adjust teaching and to enhance students' learning	to improve the teaching quality and provide students with feedback	√
		ongoing, throughout the semester	√
Form-focused assessments (FonFs, focus on forms)	assessments that divide the language into discrete elements and emphasize linguistic features	to subdivide the language into different features separately, such as grammar and vocabulary	√
		to measure linguistic features	√
Meaning-focused assessments (FonF, focus on meanings)	assessments that focus more on the comprehension of the meaning than linguistic features	to evaluate students' understanding of the meaning	√
		to pay attention to students' ability to communicate by using the target language	√

Note: Definitions and features are based on Carreira and Chik (2018) and Shintani (2013).

5 Discussion

The study examines the applicability of ChatGPT in a differentiated classroom from four aspects: assessment, content and materials, process and product. Results reveal the ChatGPT's potential and capability in contributing to effective differentiated instruction. Firstly, the moderately significant

correlation between the original HSK test scores and those evaluated by ChatGPT shows its potential for preliminary analysis of Chinese learners' linguistic performances, encompassing vocabulary, grammatical structures and nonlinguistic knowledge like the structure of the argumentation. While self-reports can be employed to evaluate students' language learning backgrounds - be it heritage or non-heritage, interests, motivation - it may be time-intensive to create, administer and grade proficiency tests. A practical method of assessment that could alleviate teachers' workload involves a written task on a uniform topic, or a topic tailored to students' abilities and interests, such as a self-introduction or description of their previous educational experience (Burns & Richards, 2012). Creating a written task to be administered in or outside of the classroom is not challenging. They can be subsequently evaluated by ChatGPT. The correlation coefficient suggests that ChatGPT has potential as a tool for grading students' writings. However, it is important to note that the strength of the relationship is not very strong. Therefore, although ChatGPT can provide a preliminary information on students' writing performance, teacher should subsequently cross-check its assessments to ensure accuracy.

Some other limitations of using ChatGPT as a grading tool should be noted. It cannot identify misspelling mistakes (such as missing a stroke in a character) as handwritten work needs to be inputted digitally before ChatGPT can grade it. In addition, within the confines of our trials where grammar correction was not particularly underscored, feedback from ChatGPT seemed to emphasize more on the content rather than the grammatical aspect of a written piece. It is worth mentioning that altering the prompts to explicitly instruct ChatGPT to focus on grammar may yield different results. Based on our trials, ChatGPT seemed to tolerate some grammatical mistakes when the semantic meaning was clear. This does not mean that ChatGPT cannot detect grammatical mistakes, its inclination towards content analysis is more pronounced without being specifically instructed to focus on grammar. For example, it can identify an incomplete sentence in a paragraph, such as " 我想跟我父亲一样, 在日本许多 ", and offer feedback such as "the structure and completeness of the writing needs to be improved". While ChatGPT may not match the expertise of an experienced rater, it is still a useful tool in assessing students' language abilities and revealing their strengths and weaknesses. As a moderately reliable rater, ChatGPT enables teachers to promptly assess students' performances, aiding in adjusting their teaching plans.

Secondly, ChatGPT can generate materials of different genres, catering to levels from A1 to C2 based on CEFR. In general, data from the L2C Rater and CTAP both demonstrate the increasing complexity of texts generated by ChatGPT as the levels rise. Within the constraints of our analyses, we can argue that while texts generated by ChatGPT often exhibit an alignment with CEFR' levels, as evidenced by its general trend of increasing scores with more advanced levels, there are exceptions particularly at higher levels. The analysis of text complexity suggests that ChatGPT has the capability to generate teaching content and materials across different difficulty levels as outlined by the CEFR. Teachers can use ChatGPT to differentiate materials by providing ChatGPT with appropriate prompts. It offers teachers the flexibility to adjust content and materials to cater to students with various proficiency levels, differentiating difficulty across diverse aspects, such as character, vocabulary, and sentence. Consequently, even when focusing on one topic, they can employ ChatGPT to generate texts of varying difficulty while maintaining the same topic. In addition to difficulty gradations, ChatGPT can also create content and materials in different genres, including description, narrative, instruction, explanation, and argumentation. Distinct discourse features are evident for each genre. For instance, an argumentation text often contains a central statement along with supporting evidence, while an instruction text would stress directives and information content. Given the potential for discrepancies, it is essential for teachers to monitoring the generated materials.

Thirdly, to differentiate teaching processes, ChatGPT can produce various tasks, including focused and unfocused tasks, tailored to the topic and language level specified in prompts. If a list of tasks has been designed, teachers can further request ChatGPT to choose specific tasks and justify the choices. ChatGPT can provide language support, such as vocabulary lists based on designated reading materials, assisting learners with limited vocabulary knowledge. Tasks and activities are central to teaching

processes, enabling students to understand information, ideas and skills effectively (Theisen, 2002). However, preparing various tasks and activities can be time-consuming, given the need to account for individual student differences, diverse task forms, objectives, evaluations, and so on. The evaluation of the tasks generated by ChatGPT reveals that ChatGPT basically adheres to the stipulations detailed in the prompt, such as the topic of the task and students' language proficiency. The topic of the task aligns well with its forms. For instance, for the topic of climate change, ChatGPT created such task forms as news analysis, policy making, filming a documentary showing the impact of climate change and solutions. These forms are fitting for a topic like climate change, as the topic delves into facts, opinions, standpoints, and arguments. In addition to the topic, the structure and complexity of tasks are tied to learners' proficiency. For beginning learners who may find tasks such as debates challenging were provided with such activities as comprehension of short sentences and passages and building vocabulary on climate change. In contrast, for intermediate and advanced learners, ChatGPT created a variety of more complex activities, such as storytelling, policy analysis, and filming a documentary.

To make tasks suitable for learners of different strengths and weaknesses, teachers may need to provide some additional language support, such as vocabulary lists, and offer mandatory and optional assignments. The result reveals that ChatGPT is able to generate vocabulary lists based on different reading materials. In our study, without specific guidance, ChatGPT generated a list of twenty-eight words from a simulated text on climate change. Chinese characters, pinyin as well as the English meaning were all included in the list. While such a list would still require the evaluation by teachers, the assistance from ChatGPT considerably reduces the teachers' workload. ChatGPT is also capable of designing and recommending mandatory and optional assignments. For instance, when presented with three assignments, ChatGPT can illustrate the advantages of one as an optional assignment and specify the skills it hones, assisting teachers in determining its suitability.

While ChatGPT exhibits abilities in creating various tasks, educators should be aware that the quality of the generated tasks largely hinges on the provided prompts. Some recommendations for writing prompts are made here. First, specificity is crucial. If only the name of a task type is provided without a detailed description, ChatGPT might generate a task without the intended characteristics. For example, when prompted with "designing a focused task", ChatGPT has simulated a task that aims at improving leaners' ability to comprehend the topic and understand the cultural information, which is more concerned with an unfocused task rather than a focused task. To address this, prompts should provide clear and detailed descriptions of the desired task type.

Second, requests for evaluating learners' performances should be clear. At the end of each task outline, ChatGPT would provide the criteria for evaluating learners' performances, corresponding to the objective of the task, such as the mastery of the sentence structure, and understanding the topic of the task. The evaluation criteria provided by ChatGPT can be general and abstract, requiring more refined prompts. For example, in evaluating the debate and discussion, the criteria are similar, encompassing aspects like linguistic accuracy, sentence complexity, pronunciation, and engagement. To align the evaluation closely with the teaching goals, teacher may need to revise the provided criteria or input more specific prompts. Finally, information on the time limit should be included in the prompts for designing tasks. In the absence of a specified time frame, ChatGPT can develop tasks that might extend beyond a single lesson, including both preparation and evaluation. Therefore, it is crucial to include a time limit in the prompt so that ChatGPT can design a task fitting in the designed duration.

Fourthly, in terms of product differentiation, ChatGPT can generate various assessments or assignments even without detailed feature inputs. ChatGPT can discern features of assignments/ assessments tested in this study, tailoring them to cater to students' needs, interests, and proficiency levels. The analysis of assessments that are generated by the two types of prompts, summative or formative, reveals that ChatGPT focuses on the key characteristic of each type, such as the ongoing assignment for incremental evaluation or the final measure of what learners have learnt and can do at the end of the course. In formative assessment, the emphasis is on teacher feedback, guiding further

instruction, and informing students about areas of improvement, whereas summative assessment primarily focuses on the final learning outcome and the use of rubrics for grading. For example, summative assessments created by ChatGPT includes oral and poster presentations, research papers, and grammar and vocabulary tests, all stressing the learning product. In contrast, formative assessments generated by ChatGPT involve writing a series of diaries and conducting regular discussions, prioritizing continuous learning and regular feedback to incrementally enhance students' language ability. ChatGPT also factors in learners' proficiency by employing different forms. For learners at intermediate and advanced levels, ChatGPT tends to use the forms that assess their integrative skills, such as oral presentations, research paper writings, and debates. In contrast, ChatGPT produces more fundamental assessments like quizzes and poster presentations for learners at the beginning level. It is worth noting that although ChatGPT seems to tailor assessment formats to different proficiency levels, it may still produce formats less typical for certain levels, such as generating a grammar test for advanced learners. In case that ChatGPT only provides the description of a quiz or test, additional detailed prompts would be necessary to obtain specific sample questions.

ChatGPT can also discern features of the form-focused and meaning-focused assignments without requiring detailed descriptions in the prompts. Examples of meaning-focused assessments are oral debates and picture storytelling, emphasizing comprehension and production for meaningful communication. In contrast, form-focused assignments delve into discrete elements of the language, grammar and vocabulary tests and examinations on sentence structures. Assignment differentiation is evident based on learners' proficiency. Beginning learners were provided with activities like picture storytelling, utilizing their known words and sentence structures to show their understanding of the topic.

6 Conclusion

Teachers should play an important and active role in the application of ChatGPT in language learning. They are accountable for understanding students' differences, creating materials, orchestrating tasks and evaluations. Following ChatGPT's grading of students' work, teachers should review the grading, pinpointing mistakes that ChatGPT might miss. When presented with a range of materials, tasks and assessments generated by ChatGPT, teachers need to assess the quality of materials, feasibility of tasks, and pertinence of assessments. Although differentiated instruction aims at addressing students' differences, it is impractical to account for each student's characteristics. It is teachers' responsibility to balance students' differences and similarities and incorporate them in the instruction with the help of ChatGPT. While choosing from many options provided by ChatGPT may seem challenging, teachers should perceive ChatGPT as a useful tool that not only saves their time but also aids in the development of diverse students in the same class.

It is worth noting that results and discussion of the study are based on the trial tests conducted with ChatGPT. As an AI model, ChatGPT is highly responsive to prompts and variations in prompts can lead to diverse responses. Therefore, fully understanding the efficacy of using ChatGPT for differentiated instruction needs more investigation, especially concerning the influence of different prompts.

Notes

1. Formative and summative assessments, the former emphasizing the learning process and the latter emphasizing the end result, can collaboratively assist students in receiving feedback and reflecting on their learning (Nieminen & Tuohilampi, 2020).
2. The instruction that focuses on forms tends to divide the language into discrete elements, such as words and grammar, and to teach them one by one in a linear way, whereas for focus on form instruction, the focus of teaching is the meaning rather than form (Shintani, 2013).

References

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57,* 100745–100750. https://doi.org/10.1016/j.asw.2023.100745

Bi, M., Struyven, K., & Zhu, C. (2023). Variables that influence teachers' practice of differentiated instruction in Chinese classrooms: A study from teachers' perspectives. *Frontiers in Psychology, 14,* 1124259–1124259. https://doi.org/10.3389/fpsyg.2023.1124259

Birnie, B. F. (2015). Making the case for differentiation. *The Clearing House, 88*(2), 62–65. http://www.jstor.org/stable/43999575

Blaz, D. (2006). *Differentiated instruction: A guide for foreign language teachers.* Eye On Education.

Blaz, D. (2016). *Differentiated instruction: A guide for world language teachers* (2nd Ed.). Routledge.

Bondie, R., & Zusho, A. (2018). *Differentiated instruction made practical: Engaging the extremes through classroom routines.* Routledge.

Burns, A., & Richards, J. (2012). *The Cambridge guide to pedagogy and practice in second language teaching.* Cambridge University Press.

Cai, W. (2023a). *ChatGPT can be powerful tool for language learning.* University Affairs.

Cai, W. (2023b). Learning and teaching Chinese in the ChatGPT context. *Language Teaching and Linguistic Studies,* (4), 13–23.

Carreira, M., & Chik, C. H. (2018). *Differentiated teaching: A primer for heritage and mixed classes. In P. Kim (Ed.), The Routledge handbook of Spanish as a heritage language* (pp.359–374). New York: Taylor & Francis Group.

Chien, C-W. (2012). Differentiated instruction in an elementary school EFL classroom. *TESOL Journal, 3*(2), 280–291.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford University Press.

Fisher, D., & Frey, N. (2008). *Better learning through structured Teaching: A framework for the gradual release of responsibility.* Association for Supervision & Curriculum Development.

Gibbs, K. (2023). Voices in practice: Challenges to implementing differentiated instruction by teachers and school leaders in an Australian mainstream secondary school. *Australian Educational Researcher, 50*(4), 1217–1232. https://doi.org/10.1007/s13384-022-00551-2

Gregory, G. H., & Chapman, C. (2013). *Differentiated instructional strategies: One size doesn't fit all (3rd Ed.)*. Corwin.

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation, 5*(1), 37–45.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 102274–102282.

Kohnke,. L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation, 17*(1/2), 214–226. https://doi.org/10.1504/IJMLO.2023.10053355

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 3368822311628–550. https://doi.org/10.1177/00336882231162868

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *TESL-EJ, 27*(3), 1–19.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Moosa, V., & Shareefa, M. (2019). The impact of teachers' experience and qualification on efficacy, knowledge and implementation of differentiated instruction. *International Journal of Instruction, 12*(2), 587–604. https://doi.org/10.29333/iji.2019.12237a

Natova, I. (2021). Estimating CEFR reading comprehension text complexity. *The Language Learning Journal, 49*(6), 699–710.

Nieminen, J. H., & Tuohilampi, L. (2020). 'Finally studying for myself' – examining student agency in summative and formative self-assessment models. *Assessment & Evaluation in Higher Education, 45*(7), 1031–1045.

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning & Teaching, 6*(1), 1–38. https://doi.org/10.37074/jalt.2023.6.1.23

Shi, Y. J., & Hua, G. D. (2007). The discussion of differentiated instruction and equity of education. *Education Study, 28*(1), 36–40.

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginning-level learners. *TESOL Quarterly, 47*(1), 36–62.

Theisen, T. (2002). *Differentiated instruction in the foreign language classroom: Meeting the diverse needs of all learners.* The Communique, (6), 1–8.

Tileston, D. W. (2004). *What every teacher should know about diverse learners.* Corwin Press.

Tomlinson, C. A. (2003). *Fulfilling the promise of the differentiated classroom: Strategies and tools for responsive teaching.* Association for Supervision & Curriculum Development.

Tomlinson, C. A. (2005). *How to differentiate instruction in mixed-ability classrooms.* Association for Supervision & Curriculum Development.

Tomlinson, C. A., & Allan, S. D. (2000). *Leadership for differentiating schools and classrooms.* Association for Supervision & Curriculum Development.

Tomlinson, C. A., & Eidson, C. C. (2003). *Differentiation in practice: A resource guide for differentiating curriculum, Grades K-5.* Association for Supervision & Curriculum Development.

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 1–25. https://doi.org/10.1007/s10639-023-11742-4

Qiuchen Li is a PhD candidate majoring in applied linguistics at the University of Calgary, Canada. Her research interests include second language acquisition, teaching Chinese as a second language, second language listening, vocabulary assessment and differentiated instruction.

Jiafan Zhang, MA in Applied Linguistics of University of Calgary. Research interests include second language acquisition, Chinese listening, Chinese heritage and non-heritage Leaners, and differentiated instruction.

Dr. Wei Cai is a Professor of Chinese in the School of Languages, Linguistics, Literatures, and Cultures at the University of Calgary, Canada. Her research interests include second language listening and second language acquisition. She was awarded the Robert Roy Award by the Canadian Association of Second Language Teachers (CASLT) in 2023. Her new book, *Teaching and Researching Chinese Second Language Listening,* was published in 2022 by Routledge (Taylor & Francis Group).

国际汉语教学学报（2024）
第 05 卷 · 第 01 期 74-89 https://doi.org/10.46451/ijclt.20240106

ChatGPT 在差异化教学中的应用

李秋晨
张佳帆
蔡薇
卡尔加里大学，加拿大

摘要
本文研究了使用 ChatGPT 辅助差异化教学的可能性，重点研究其评估汉语学习者等级，针对不同水平的学生生成不同文本、设计不同的教学任务以及测试的可操作性。通过随机方式选出 45 篇 HSK 作文文本，进一步计算文本原始分数与 ChatGPT 生成的分数之间的相关性。此外，针对不同水平的学习者，ChatGPT 产出了 30 篇不同类型的教学文本，从而检测其教学文本生成能力。最后，本文检验了 ChatGPT 输出不同任务及测试的能力。

研究结果显示 HSK 作文文本原始分数与 ChatGPT 生成分数之间存在显著的相关关系，反映了其评估汉语学习者水平的可能性。它能够针对不同水平的学生输出不同类型教学文本，难度等级与 CEFR 要求一致。如果输入清楚具体的指令，ChatGPT 可以灵活地输出教学任务和测试。ChatGPT 在差异化教学中的应用还需要更多深入的研究。

关键词
ChatGPT, 差异化教学 , 汉语学习者 , 语言能力

李秋晨：加拿大卡尔加里大学博士研究生，研究方向为应用语言学，研究兴趣包括二语习得、汉语国际教育、二语听力、词汇测试、差异化教学。

张佳帆：加拿大卡尔加里大学应用语言学硕士。研究兴趣为二语习得，汉语听力，华裔与非华裔汉语教学，以及差异化教学。

蔡薇：加拿大卡尔加里大学语言、语言学、文学及文化系教授，研究兴趣包括第二语言听力与第二语言习得。2023 年荣获加拿大第二语言教师学会颁发的 Robert Roy 奖。新书《汉语第二语言听力教学与研究》由 Routledge (Taylor & Francis Group) 出版社出版。