International Journal of TESOL Studies (2026)
261105, 1-16 https://doi.org/10.58304/ijts.261105

Article

TESOL Practitioners' Perceptions of Using ChatGPT to Design Communicative Tasks for Task-based Language Teaching

Martin East
University of Auckland, New Zealand

Received: 29 September 2026/Received in revised form: 27 October 2026/Accepted: 28 October 2026/Available online: 2 November 2026

Abstract
Task-based language teaching (TBLT) is increasingly advocated in a range of contexts across the globe as an effective learner-centred and experiential alternative to more traditional teacher-led and instructivist approaches. However, its implementation in classrooms has not always been straightforward. In practice, teachers often struggle not only with understanding what a task is for the purposes of TBLT but also with how tasks fit into the broader learner-centred and experiential pedagogical approach that informs TBLT. Additionally, time constraints and workload pressures have been identified as significant barriers to effective TBLT implementation. In the context of teacher education, this paper considers to what extent ChatGPT as an open access generative artificial intelligence application can support TESOL practitioners with not only generating tasks that are appropriately aligned to TBLT theory but also streamlining the lesson and resource planning and preparation process. The present study investigated what 39 participants in a dedicated postgraduate course focused on TBLT thought of using ChatGPT to design a task, with associated pre-task and post-task activities, for potential classroom use. It draws on data from individual reflective presentations given at the end of the course. Six themes were constructed from a reflexive thematic analysis – three benefits and three limitations of using ChatGPT for task creation purposes. In light of perceived advantages and disadvantages, participants saw using ChatGPT for task generation as a 'yes … but' proposition. The findings raise key issues for teachers and teacher educators to be mindful of when using tools like ChatGPT to design tasks.

Keywords
Task-based language teaching (TBLT), task design, generative artificial intelligence (GenAI), ChatGPT, practitioners' perspectives

1 Introduction

The approach to language pedagogy known as task-based language teaching (TBLT) is increasingly advocated in a range of contexts across the globe as an effective learner-centred and experiential

Email: m.east@auckland.ac.nz

Online First View

alternative to more traditional teacher-led and instructivist approaches. The essential point of differentiation between TBLT and other approaches is the task itself. Ellis (2024), for example, described tasks within TBLT as "the unit around which individual lessons are built," meaning that lesson planning "begins with a task" (p. 6). It is therefore important for teachers who wish to implement a task-oriented approach to start by developing a clear understanding of the task construct.

In practice, however, teachers often struggle not only with understanding what a task is for the purposes of TBLT but also with how tasks fit into the broader learner-centred and experiential pedagogical approach that informs TBLT (e.g., Bryfonski, 2022; Bryfonski & Pineault, 2026). Additionally, when it comes to the adoption of innovative teaching strategies such as TBLT, time constraints and workload pressures have been identified as significant barriers to implementation (Belloula, 2025; Sun, 2026). Teacher education is regarded as one means through which the challenges of putting TBLT into practice can be mediated (e.g., Bryfonski, 2022; East, 2022; Van den Branden, 2006). A key question becomes how teacher education can be used to support teachers with implementing TBLT and tasks more effectively.

This article considers to what extent ChatGPT as an open access generative artificial intelligence (GenAI) application can support teachers with designing tasks in the context of teacher education. GenAI has been identified as one means of streamlining lesson and resource planning and preparation aligned with appropriate theoretical and methodological approaches, leading to a significant reduction in the time that teachers need to spend searching for or creating suitable resources (Belloula, 2025; Kerr & Kim, 2025; Moorhouse, 2024). The present study investigated how participants in a dedicated postgraduate course focused on TBLT responded to using ChatGPT to design tasks for potential classroom use. It draws on data from individual reflective presentations given at the end of the course, and presents and discusses what the participants perceived as the benefits and limitations of using ChatGPT for task creation purposes.

2 Literature Review

2.1 The theoretical and operational bases for TBLT

The theoretical case for TBLT has been argued from both a second language acquisition (SLA) and an educational perspective (Long, 2015). The SLA standpoint is supported by Long's suggestion that learning and SLA will occur when learners completing a communicative task engage with the cognitive demands placed on them as they negotiate meaning and process feedback during task completion – the cognitive-interactionist position (e.g., Long, 1981, 1996). From an educational standpoint, the role of the teacher shifts from that of the expert who instructs and transfers knowledge to learners to that of the facilitator who oversees learners as they construct knowledge for themselves – the sociocultural-interactionist position (e.g., Vygotsky, 1978).

TBLT's departure from more established pedagogical practices has led to the perception that TBLT is "still a relatively recent innovation" (Long, 2016, p. 28). Also, TBLT's strong learner-centred and experiential orientation has been challenged as the take-up of TBLT ideas has diversified to a wide range of countries and educational contexts. It is clear that TBLT is now implemented in a range of ways that may differ from its original conceptualisation, including in ways that admit a stronger instructional role for teachers (e.g., Lu et al., 2026).

A series of articles published in the *International Journal of TESOL Studies* (Vol. 6, Issue 4) explored different ways of putting TBLT into practice as invited contributors responded to a position paper by Rod Ellis. Ellis (2024) framed the different approaches as task-based (i.e., TBLT which is built on a learner-centred and experiential orientation) and task-supported (i.e., TSLT which foregrounds more teacher-led and instructivist elements). He went on to consider how the strengths of both approaches

Online First View

may be realised concurrently. Bui (2024) challenged the notion of whether explicit language instruction within a TSLT approach was beneficial or necessary, arguing that the more implicit approach advocated within TBLT was adequate to address the development of grammatical competence. Broadening out a dichotomous understanding of TBLT, East (2024) spoke of a continuum of practice from stronger forms of TBLT at the more learner-centred and experiential end to weaker forms (or TSLT) at the more teacher-led and instructivist end. Boers (2024) revisited the well-established Presentation-Practice-Production mode of classroom teaching and learning, reframing it as a flexible means of attending to both forms and meaning.

In essence, contributors to the volume supported Ellis's (2024) assertion that TBLT as an approach is not 'monolithic', that is, not fixed, uniform, and rigid (Boers, 2024; Bui, 2024; Chen & Dan, 2024; East, 2024; Teng, 2024; Zhu, 2024). Putting aside the delivery mode or 'flavour' of TBLT, a fundamental aspect of delivery is the task itself. That is, a task-based lesson might involve three phases (pre-task/task/post-task) which allow considerable room for teachers to tailor approaches to local contexts, but the *task* is the central vehicle for learner-centred and experiential learning, and the means through which ownership is placed on the learners themselves to experiment with language.

Although there is debate among TBLT scholars about exactly how a task should be defined for the purposes of TBLT (see Bryfonski & Pineaut, 2026, for a very recent discussion), the construct of task within TBLT has largely been understood from a particular theoretical perspective that includes a number of crucial elements. Ellis presented his own definition as "a set of criteria that could be applied to specific workplans to determine whether they can be considered tasks" (Ellis, 2024, p. 2):

1. The primary focus is on meaning.
2. There is some kind of gap.
3. Learners rely mainly on their own linguistic and non-linguistic resources.
4. There is a clearly defined communicative outcome.

Bui (2024) referred to Ellis's four-fold task definition as "the gold standard in the field" (p. 16). Reflecting this definition, Willis and Willis (2007) suggested that task-likeness could be determined by the extent to which a given task could be said to align positively with a set of six questions:

1. Does the activity engage learners' interest?
2. Is there a primary focus on meaning?
3. Is there an outcome?
4. Is success judged in terms of outcome?
5. Is completion a priority?
6. Does the activity relate to real world activities? (p. 13).

2.2 The teacher's role in task implementation

Despite theoretical definitions such as those presented by Ellis (e.g., 2024) or Willis and Willis (2007), for many teachers, *task* remains "a somewhat fuzzy construct" (Bui, 2024, p. 16; Ellis, 2024, p. 4) which can lead to teacher uncertainty about what a task actually is for the purposes of TBLT (Boers, 2024; Chen & Dan, 2024; Teng, 2024; Zhu, 2024). However, teachers are unquestionably determinants in the effective implementation of TBLT (Long, 2016; Van den Branden, 2016), pointing to the need for carefully designed teacher education programmes, for both beginning (pre-service) and more experienced (in-service) teachers (Ellis, 2017). Teacher education may lead, for example, to enhanced understanding of TBLT and tasks as teachers are supported with applying the theoretical characteristics of a task to a given task planning exercise (Ellis, 2015; Erlam, 2016; Han, 2018).

Online First View

Studies into teacher education and TBLT have also revealed that time constraints can be a particular barrier to the successful implementation of TBLT, due largely to the time needed to develop effective and theory-aligned tasks and the lack of 'go to' resources. Teachers, who frequently face a range of demands both within and beyond the classroom, report that they often simply do not have sufficient time to devise and evaluate tasks (East, 2022; Hasnain & Halder, 2023; Pham & Nguyen, 2018). This means that, in practice, they may well (at least occasionally) rely on prefabricated materials or textbook exercises. This may undermine the level of success with implementing TBLT. One avenue of exploration is on how GenAI might be used to support teachers in their work, since GenAI would seem to offer "a promising solution to alleviate the workload and enable teachers to focus on student-centered learning" (Belloula, 2025, p. 265).

2.3 Using GenAI to support teacher planning

Over recent years, it has become impossible to avoid the term 'generative artificial intelligence' or GenAI. Large Language Models (LLMs) represent a particularly powerful type of GenAI because they are designed to simulate how humans organise language and process and interpret information. This enables the LLM to understand human input and respond to it in a natural way. Generative Pretrained Transformer (GPT) models represent a type of LLM that has increasingly become embedded into our lives in many arenas. First released in November 2022, ChatGPT, developed by OpenAI, is perhaps the most well-known and accessible (although its use is restricted in some contexts). Tools such as ChatGPT can accomplish a very wide variety of activities when given straightforward language input (Bonner et al., 2023; Cain, 2024).

Since GenAI tools such as ChatGPT are now becoming ubiquitous, it is important to find ways to integrate their use into the teaching and learning process. It is already apparent that teachers are using GenAI tools including ChatGPT to help them plan lessons (Moorhouse, 2024), and lesson and resource preparation is one area where the benefits of AI are arguably strong. Due to the vast range of data that ChatGPT can call on, it is also capable of aligning its output to a stated theoretical framework if instructed to do so (Kerr & Kim, 2025). It is, nonetheless, far from a perfect tool (at present at least). It can make mistakes (sometimes referred to as hallucinations) and its output requires scrutiny (Belloula, 2025; Dornburg & Davin, 2025).

2.4 Studies into GenAI use and teachers' lesson planning

Although there is potential for GenAI to support teachers with creating tasks, teachers' use of GenAI for task design within a TBLT framework is a thus far underexplored research area. However, beyond the specific realm of TBLT, several very recently published studies have explored the use of GenAI tools in the context of teacher education and teachers' lesson planning. The findings of these studies bring out some of the benefits, but also the limitations.

Belloula (2025) investigated how secondary school teachers of a range of subjects in Qatar (n = 10) integrated AI use into their lesson planning processes, drawing on surveys, interviews, and observations derived from anonymised examples of AI-generated lesson plans. On the positive side, teachers' reporting suggested that, on average, AI tools reduced time spent on lesson planning by an estimated 40%. AI also created lessons that were perceived as highly engaging, contributing to enhanced student participation. Nonetheless, it seemed that, at times, the AI-generated output did not fully align with the direct cultural context of the learners, meaning that the generated plans required a level of adaptation to the context. Belloula concluded that, whilst GenAI might operate as a useful assistant, its use "should *complement*, not *replace*, the role of teachers" (p. 276, my emphases). It was necessary for a balance to be maintained between AI use and human interaction, due to the teacher's central relational and oversight roles.

Online First View

Focusing specifically on language teaching, Moorhouse et al. (2025) explored pre-service language teachers' GenAI-assisted lesson planning in the context of a postgraduate course – *Grammar for Teaching* – within the Master of Education in English Language Teaching (MEd: ELT) at a Hong Kong university. The first author was the lecturer for the course. One course assignment required the students to work in groups of three or four to create a single 45-minute grammar-focused lesson plan. Participants were then tasked, individually, with using LLMs to refine the plans after receiving formative feedback from the lecturer.

The dataset (Assignments from 25 students) comprised four elements: (1) the modified lesson plan; (2) the log of LLM use; (3) a description of the changes made and the perceived benefit of using the LLMs; and (4) a reflection on the use of the LLMs in the lesson planning improvement task. The findings revealed "a complex and dynamic interactional sequence with pre-service teachers engaged in a cyclical process of prompting, evaluation, and refinement" (p. 17). That is, the prompt will generate output from the AI agent, but that output must be scrutinised and potentially modified by teachers as human agents. Mirroring Belloula's (2025) conclusion, Moorhouse et al. (2025) noted that LLMs "can *assist* but are not *responsible* for the content generated" (p. 3, my emphases).

The benefits and limitations of GenAI were also identified by Kerr and Kim (2025), who investigated how pre-service English as a Foreign Language (EFL) teachers of middle school students used GenAI for lesson planning. The participants (n = 8) were enrolled in a digital literacy course at a university in South Korea. Data were collected in the context of three 75-minute classes within a 16-week semester that included: an introduction to GenAI prompting techniques (Session 1); designing a 50-minute lesson with the help of GenAI (Session 2); reflections on participants' use of GenAI (Session 3). Semi-structured interviews were an additional data source, conducted during Session 3. These were audio-recorded and subsequently transcribed. The data were analysed thematically.

Findings indicated that participants used GenAI to help with topic selection, materials creation, lesson organisation, and language checking, thereby providing "efficiency and creative support" (Kerr & Kim, 2025, p. 113) and enabling the beginning teachers to pay more attention to pedagogical decisions. The lesson plans did generally adhere to the communicative orientation to language pedagogy appropriate to the target EFL classes. However, some plans showed "lack of depth in lesson content, insufficient support for advanced teaching methodologies, and occasional inaccuracies" (pp. 113-114). Kerr and Kim commented that, although these limitations may also have been present in plans the teachers may have developed independently, it seemed that GenAI was not making a contribution in these areas. They concluded that it was important not only to help teachers craft well-designed prompts as a "foundational skill for maximizing the utility of generative AI tools" (p. 112) but also to encourage critical evaluation of the AI-generated outputs, including identifying potential biases or inaccuracies.

Focusing on the prompt itself, Dornburg and Davin (2025) explored how prompt variations might influence ChatGPT generated lesson plans for language teaching. Their study utilised 'zero-shot prompting' (Kojima et al., 2022), a prompting method that inputs a single prompt and does not undertake subsequent interaction with the tool. They designed five increasingly explicit prompts. They also input each prompt ten times to determine variability in outputs. They then scored the 50 available outputs for the presence or absence of specific lesson plan criteria as given to pre-service language teachers in a teacher education programme at an American university.

It was found that ChatGPT was generally capable of creating lesson plans aligned with the criteria, indicating that zero-shot approaches can generate valid target outputs. Interestingly, increasing the detail and specifications of the prompt did not lead to an increase in output score (i.e., better or more aligned lesson plans). On the other hand, there was considerable variability in quality across the plans, even given the same single prompt. In some cases, the plans favoured more teacher-led and behaviourist-informed approaches, appearing to overlook a more learner-centred and experiential orientation. Dornburg and Davin (2025) surmised that because GenAI models derive their responses

Online First View

from "mathematical associations between words that form the basis of each output" (p. 343) – that is, they do not themselves critically evaluate their own output – the data they draw on may over-represent more traditional or embedded theoretical perspectives. Whilst their study emphasised the benefits of "a *synergistic*, not *antagonistic*, relationship between this emerging technology and language educators" (p. 344, my emphases), recognising potential misalignment with currently espoused theories of language teaching and learning represented "an urgent need" (p. 344).

The studies reviewed above focused on teachers' use of GenAI tools in lesson planning and preparation outside of the specific context of TBLT. A conceptual paper by Sun (2026) directly addressed the intersection between GenAI, TBLT, and the teacher's role in task design.[1] Her paper articulated several of the issues identified in prior studies. Sun proposed that GenAI could support the speedy generation of theoretically-aligned tasks for a range of proficiency levels, thereby saving teachers considerable time and making GenAI "an ideal companion to TBLT" (p. 423). She argued nonetheless that tasks created by GenAI may at times be insufficiently aligned to the learning goals of a particular context or may introduce or reinforce cultural bias. Consequently, its use did not diminish the crucial role of teachers, which shifted to evaluating and scrutinising the output to ensure adequate alignment with teaching goals and TBLT principles.

Sun's paper makes an important contribution to arguments regarding the benefits and limitations of using GenAI in the specific TBLT-planning context. However, because her paper is conceptual, no empirical evidence is provided to support her claims. Taking into account prior studies alongside Sun's (2026) conceptual paper, the present study investigated practitioners' perspectives on the use of GenAI for task generation within a postgraduate course that contributes directly to TBLT-oriented teacher education. The following research question is addressed: what do the course participants perceive as the benefits and limitations of using ChatGPT to design a communicative task for task-based language teaching?

3 Methodology

3.1 Participants

Participants were drawn from the cohort of students completing a postgraduate taught course delivered over 12 weeks in the second part of 2025 at the University of Auckland, New Zealand – *Task-Based Language Teaching*. Students taking the course were enrolled in one of three postgraduate programmes – the Masters in Teaching English to Speakers of Other Languages (MTESOL); the MA in Applied Linguistics; and the Postgraduate Diploma in Language Teaching. Although the course was offered concurrently across all three programmes, the differentiating feature of the sub-cohorts was that those enrolled in the MTESOL are required to enter the programme with the equivalent of at least two years of relevant professional experience. The qualification is therefore viewed as a professional development opportunity for those who would likely return to teaching after its completion. The Diploma is also viewed as essentially for professional development purposes, although there is no teaching pre-requisite. The MA is more academically-oriented and no prior teaching experience is presumed (although some students do have prior experience).

The present study formed part of a broader project that drew on course participants' assessed coursework as data sources that documented participants' developing understanding and application of TBLT. A total of 56 students were enrolled in the course, of whom 66% of the cohort (37 students) were in the MTESOL programme. A total of 39 students gave informed written consent to take part in the project, of whom 26 (67%) were MTESOL students. Their self-reported prior professional experience ranged from nothing (or not stated) to (in two cases) 20+ years. The mean number of years of experience was 3.8. The majority of consenting participants were from China (79%) and a larger majority (85%)

Online First View

were female. Table 1 aligns the allocated participant numbers (Participant #1, etc.) against several of these variables.

Table 1.
Participant Details

Participant No.	Programme	Nationality	Reported prior teaching experience
#1 - #12	MTESOL	Chinese	0 – 5 years
#13 - #20			5+ years
#21		Other	0 – 5 years
#22 - #26			5+ years
#27 - #28	Postgraduate Diploma	Chinese	0 – 5 years
#29 - #37	MA in Applied Linguistics	Chinese	0 – 5 years
#38 - #39		Other	0 – 5 years

The researcher taught the course. This dual role (lecturer and researcher) was not seen as one that would affect the data collection or the participants' interaction with the course because data collection, which drew exclusively on participants' coursework, was embedded in the regular course delivery and data were not collected wholly or primarily for research (see Kerr & Kim, 2025, for a similar approach).

3.2 Data collection

The 12-week taught course enabled an extensive exploration of the theory and practice of TBLT. A broad range of data was collected through several assessed assignments. One two-part assignment focused on task design and evaluation. In the first part, participants were required to download a task from the *TBLT Language Learning Task Bank* (Gurzynski-Weiss & IATBLT, n.d.), to evaluate the task for its task-likeness using the criteria presented by Ellis and Willis and Willis (see Section 2.1) and to suggest improvements. In the second part, participants were required to design their own task using ChatGPT. Again, they were asked to evaluate the task for its task-likeness, once more using the Ellis and Willis and Willis criteria, and, again, to suggest improvements. At the time of the study, the default model of ChatGPT was GPT-5. To ensure that all course participants were evaluating its use and effectiveness from a similar basis, participants were given clear guidance about what they should input into ChatGPT following a zero-shot prompt approach. They were to instruct ChatGPT to design a task based on task-based language teaching theory (although not specifically making reference to, e.g., Ellis or Willis & Willis), including pre-task and post-task activities, and encouraging pair or group interaction. There was discretion for participants to select the overall time limit for the task phases, the topic, and the target proficiency level, benchmarked against the Common European Framework of Reference to which they had been introduced in the course (Council of Europe, 2001, 2020) as a means to differentiate proficiency levels within tasks.

In the final week of the course, all participants attended an online conference, facilitated by Zoom, where they were required to present (as a summative assignment) a ten-minute pre-recorded video presentation on what they had learned about tasks and TBLT in the course. Participants were allocated randomly to break-out rooms of about 5 or 6 people and asked to share their presentations, one by one. During this process, the researcher and a graduate teaching assistant visited each room. A discrete section of the presentation was to be devoted to a critical evaluation of using ChatGPT for task design, including what participants perceived as the benefits and limitations of ChatGPT for this activity. For the purposes of the overall research project, the videos were subsequently transcribed for analysis. The section of the presentations focusing on ChatGPT forms the basis of the data analysed in the present study.

Online First View

3.3 Data analysis

The data were analysed drawing on a reflexive thematic analysis (TA) approach (Braun & Clarke, 2006, 2021). Differences between sub-cohorts (in terms of programme, nationality, or past experience as noted in Table 1) were not included as part of the data analysis since the main purpose of reflexive TA is not to quantify instances of themes and therefore to undertake a comparative analysis across sub-groups; rather, a theme represents an important shared meaning in the data set as constructed by the researcher from the data.[2] Its importance is therefore not dependent on how many people referred to it, and the primary goal is to construct themes across an entire dataset.

Bearing in mind that a broad range of data had been collected as part of the ongoing research project, an initial step was to extract the data pertaining to participants' reflections on GenAI use for task creation from the larger data set. Once that had been done, the researcher initially read through the extracted data to get an overall sense of the issues being raised. Next, a 'bottom-up' (inductive) coding approach was undertaken, where the issues the researcher observed, whether benefits or limitations of using ChatGPT for task creation, were coded. This led to the construction of an initial list of eight themes. The data were then revisited and the initial list was rationalised. Six themes were constructed from this process (Table 2).

Table 2

Themes Constructed from the TA

Benefits of ChatGPT	Limitations of ChatGPT
ChatGPT can …	ChatGPT output is …
• lead to planning efficiencies	• not always theoretically aligned with TBLT
• provide ideas a teacher may not have thought of	• not context-sensitive
• create appropriately differentiated tasks	• not a substitute for the teacher

4 Findings

The following sections draw on participant comments to illustrate the themes constructed from the data, starting with the perceived benefits (Section 4.1) and following with the perceived limitations (Section 4.2), leading to the conclusion that using ChatGPT should be seen as a 'yes … but' proposal (Section 4.3).

4.1 The benefits of ChatGPT

A core theme in the data set (referred to in one way or another by all but one participant) was that use of ChatGPT could lead to planning efficiencies (Theme 1). Words related to 'efficient' and 'time-saving' were widely in evidence. Typical comments included: ChatGPT was "very efficient" (#29) or operated as a "powerful efficiency tool" (#1) because it "saves teacher preparation time … thereby improving teachers' lesson preparation efficiency" (#16). Two participants referred to ChatGPT as "a game-changer" because it was seen as "incredibly efficient" (#4) and "a phenomenal starting point, a tool that can accelerate the design phase" (#2).

When participants elaborated on the efficiency dimension, this was often in relation to its support with effective task planning and preparation in line with the prompt instructions participants had been asked to use (see Section 3.2). ChatGPT could "generate diverse task ideas and format in just a few seconds" (#13), saving "a lot of time in the early design process" (#14) because of its ability to draw on "a vast database of task typologies and frameworks" (#10). In Participant #7's words, ChatGPT "generates a complete task cycle quickly, pre-task, task, and post-task, which saved plenty of time," making the tool "an

Online First View

efficient core planner that gives a solid starting point and practical suggestions" and "a helpful starting point for lesson design." Participant #35 explained, "by simply providing the topic, target, proficiency level, time frame, and requirements for activities," ChatGPT could "rapidly generate a complete lesson plan that includes pre-task, task, and post-task stages." Describing ChatGPT as "a power tool that speeds up good planning," Participant #36 noted that ChatGPT helped to "move faster from an idea to something teachable," thereby saving time by "drafting scaffolds that usually take me a long evening to make."

A connection was also made between ChatGPT's efficiency and its ability to "offer new ideas, fresh formats, and different ways to present tasks that teachers might not have thought of on their own" (#26) (Theme 2). This "encourages creativity" (#13) or enables teachers to "be more creative" (#3) because "sometimes it gives me [a] perspective I hadn't thought of before" (#13). This creativity dimension made ChatGPT "a very powerful tool for inspiration," helping to "generate ideas" (#18) or "inspire innovative instructional ideas" (#11). As Participant #14 expressed it, "once I gave ChatGPT a simple prompt, it instantly generated multiple task ideas with detailed procedures." ChatGPT then became a useful tool to "overcome creative blocks" (#14) or "the initial 'blank page' problem" (#32). Participant #4 put it like this: ChatGPT can "stimulate teacher creativity" because "sometimes, seeing an AI-generated example helps teachers think in new directions or refine their own ideas. In that case, ChatGPT can act as a kind of brainstorming partner."

Participants also perceived that ChatGPT could generate tasks suitable for different proficiency levels (Theme 3). This ability to "adapt tasks" (#14) or "adjust the difficulty of the task" (#13) or generate output that is "more challenging or less challenging to suit your class" (#20) meant that "tasks can be automatically tailored to specific language proficiency levels and learning objectives" (#11), thereby "helping all learners participate meaningfully" (#26). It also "supports differentiation" (#34), leading to tasks that are "more appropriate and effective for different learners" (#16). As Participant #25 explained:

> ChatGPT can facilitate differentiation and scaffolding. Teachers can request simplified or more complex versions of the same task to suit mixed ability classes, ensuring the activity is appropriately adapted to the full range of abilities present in the classroom. For scaffolding, tasks can be designed to form a series that gradually increases in difficulty, following models such as Willis's [1996] order of increasing complexity that begins with listing tasks and ends with creative tasks.

4.2 The limitations of ChatGPT

Although there appeared to be praise for ChatGPT due to its agility in creating tasks that could suit a range of learners, some participants questioned its consistent ability to align tasks to appropriate underlying theoretical frameworks (Theme 4). It seemed that ChatGPT does not always "truly comprehend the principles behind effective task design" (#10) or "constantly misinterprets analysis of TBLT theories, even when you spell them out" (#30), meaning that "despite inputting content and asking the task to be aligned with a certain methodological framework, there is no guarantee that the task created will in fact be aligned" (#21).

Several examples were given where "tasks generated by ChatGPT don't always follow a line with established TBLT theoretical frameworks, such as those by Ellis and Willis" and therefore "might miss key theoretical components" (#8). These limitations potentially make ChatGPT "not an insightful instructional designer" (#1). ChatGPT may:

- "generate homogeneous or pseudo tasks that look good on the surface but *lack real meaning negotiation*" (#13, my emphasis) – see Ellis's Criterion 1;
- generate tasks that "*fail to create gaps* that are creative enough for students to spend the time working to overcome" (#39, my emphasis) – see Ellis's Criterion 2;

Online First View

- suggest that teachers simply pre-teach useful expressions rather than "letting students actively use the language" (#2) – see Ellis's Criterion 3.
- "produce activities that *look* communicative," but "many are still drills … [that] lack key task features like *non-linguistic outcomes*" (#31, my emphases) – see Ellis's Criterion 4.

An allied concern was ChatGPT's lack of context sensitivity (Theme 5). That is, ChatGPT "lacks nuanced understanding of the students" (#18) or "real classroom awareness," resulting in tasks that "often sound logically perfect," but without insight into "how a task will actually play out in a real classroom" (#35). The output might reveal "a profound lack of pedagogical understanding" (#10) or "completely lacks pedagogical foresight" (#2). Essentially, ChatGPT "can't say if a task would work in a real classroom because it hasn't worked with real students before" (#27) and "lacks genuine lived experience" (#11). It therefore "often misses pedagogic fit because ChatGPT doesn't know my learners" (#33).

Several participants raised concerns about ChatGPT's lack of relevant cultural understanding, leading at times to tasks that were mismatched to the specific cultural context: "sometimes the facts are off … or the culture references don't fit our context" (#36); "sometimes the information provided may be incorrect, outdated, or contextually inappropriate" (#9); "sometimes the responses are too general or lack cultural and contextual accuracy" (#28). As Participant #26 put it, "a task that works well in one country may be irrelevant or even inappropriate in another."

Some participants placed the blame for ChatGPT's cultural and contextual inaccuracies on "biases present in its training data, leading to subtle stereotypes or unbalanced cultural perspectives" (#25). That is, "ChatGPT relies on existing data patterns so the task it generated can sound generic or fail to reflect the learner's cultural or local context" (#31). More concerningly, "because AI systems are trained mostly on English-dominant data, they may reflect Western cultural norms and even hidden biases. This can influence how topics are framed and whose voices are represented" (#15), leading to "misleading or culturally biased content as the AI might generate inaccurate or insensitive information" (#13).

Overall, it seemed that often the output from ChatGPT presented "a polished one-size-fits-all blueprint, assuming a perfectly average classroom that simply doesn't exist" (#2) or "one-size-fits-all task materials" (#30) that ignore cultural or contextual fit. This leads to tasks that are "too general and not specific to my teaching context" (#3) because ChatGPT "doesn't really know our students, like their language levels, learning goals, or feelings" (#15) or "doesn't understand the nuances of the varying proficiency levels in the class, the specific challenges that every student faces" (#37). Since ChatGPT "cannot fully consider the specific overall level and individual differences within a particular classroom" (#8) because it "lacks context awareness" (#34), it "has trouble predicting how real human students act in a classroom" (#39). Participant #20 noted, "I realised that GPT pretty much gives you the same task structure no matter what you ask it to create." This could mean that "if you overuse ChatGPT for your tasks, students might get bored, so it defeats the purpose of making learning fun by using tasks."

In light of the above identified constraints, most participants commented that ChatGPT use could not become a substitute for the teacher (Theme 6). Critical evaluation of output drawing on teachers' "professional knowledge and classroom experience" (#27) was important and necessary. It was not only "still the teacher's responsibility to ensure that the task meets pedagogical principles" (#24); it was also important to verify that "the content is culturally appropriate and inclusive for our own students" (#15). That is, teachers need to: "interpret, adapt, and humanise what AI creates to make sure that tasks are realistic, ethical, and appropriate for our learners" (#15); "verify contents carefully" (#9); and "refine and adapt the materials it generates to make sure they truly fit students' needs" (# 13).

Online First View

4.3 ChatGPT use – 'yes … but'

In a high number of instances, Theme 1 (ChatGPT can lead to planning efficiencies) and Theme 6 (ChatGPT output is not a substitute for the teacher) converged, revealing that several participants saw partnering with ChatGPT as a 'yes … but' proposition:

- "a valuable support tool … but it cannot replace the teacher's role" (#3);
- "a collaborative tool, not a replacement" (#4);
- "a powerful assistant rather than a replacement for teachers" (#9);
- "a highly effective assistant in the design of TBLT lessons, but it cannot and should not replace the teacher" (#11);
- "a very useful assistant for generating task ideas and structures, but it cannot replace teachers" (#15);
- "a supporting tool, not as a replacement for professional expertise" (#34).

As Participant #10 put it, ChatGPT was "a powerful copilot, but it should never be the pilot." Other participants elaborated on the dimensions where they perceived the 'yes … but' distinction. Participant #18 spoke of ChatGPT as "a useful assistant for task design, offering creativity and speed." The 'but' in this case was that "the teacher's judgments and experience remain central to effective TBLT implementation." For Participant #24, ChatGPT could operate as "a creative assistant" and "a valuable support tool," but it could not replace "teacher expertise in adapting materials to a real classroom." Hence, it needed to be used "critically and selectively." Participant #7 referred to ChatGPT as "a helpful starting point for lesson design." However, "it needs a teacher's judgment and classroom piloting to be effective." The assistance offered by ChatGPT may be "fantastic," but "it cannot and will not replace the teacher's critical role as a responsive and empathetic facilitator" (#2).

The conclusion reached by a number of participants was that "while ChatGPT is a powerful tool to support task creation, it's most effective when used alongside the teacher's knowledge and professional judgment" (#26). As Participant #28 put it, "I believe the future of language education lies in the collaboration between human creativity and AI intelligence."

5 Discussion

The purpose of the present study was to explore TESOL practitioners' perceptions of using ChatGPT to design communicative tasks for TBLT in the context of a postgraduate taught course. Data were drawn from transcripts of participants' summative video presentations where they were asked to share with each other in small groups their experiences with using ChatGPT, reflecting on and identifying both benefits and limitations. The participants had not been required to undertake any pre-reading in the area of GenAI use, since the primary emphasis of the course was on TBLT. Bearing this in mind, it is noteworthy that the findings align with and confirm those of previous research. Six themes were generated in the data analysis.

With regard to benefits, a dominant theme was that ChatGPT use could lead to planning efficiencies. The time-saving dimension of ChatGPT has also been highlighted previously and has been identified as a major contribution that tools like ChatGPT can make (Belloula, 2025; Moorhouse, 2024; Sun, 2026). In the words of Participant #25, it seemed that ChatGPT "directly addresses one of the main drawbacks of TBLT … its time-consuming nature" (see also East, 2022; Hasnain & Halder, 2023; Pham & Nguyen, 2018). Kerr and Kim (2025) spoke of efficiency *and* creative support, and the *creative* dimension of ChatGPT use – providing ideas a teacher may not have thought of – was in evidence in Theme 2. ChatGPT's ability to create suitably differentiated tasks (Theme 3) finds resonance in prior studies that

Online First View

speak of the ability of GenAI to align with an appropriate orientation to language pedagogy (Dornburg & Davin, 2025; Kerr & Kim, 2025).

With regard to limitations, some concerns were raised that ChatGPT output was not always theoretically aligned with TBLT principles (Theme 4). Lack of alignment with espoused contemporary theories of effective pedagogical practice is once more in evidence in other studies (Belloula, 2025; Kerr & Kim, 2025). In line with a concern raised by Dornburg and Davin (2025), there was some recognition that ChatGPT was only as good as the data it was able to draw on and its outputs might be biased as a consequence. Outputs revealed ChatGPT's occasional lack of context-sensitivity, an issue also identified, for example, by Belloula (2025). Theme 6 – that, in light of constraints, ChatGPT could not be a substitute for the teacher – finds resonance in previous work (Belloula, 2025; Dornburg & Davin, 2025; Kerr & Kim, 2025; Moorhouse et al., 2025).

One contradictory finding was that ChatGPT-generated tasks were suitably differentiated for a range of proficiency levels and, conversely, did not differentiate between levels, leading to 'one-size-fits-all' outputs. Also contradictory was that, although the benefits of planning efficiencies were strongly in evidence across the data set, there was also reference to the fact that "while the ability to refine and redefine prompts is powerful, the process of continual negotiation and correction can ultimately consume a significant amount of a teacher's time and patience" (#25), meaning that ChatGPT use "could be time consuming if a teacher is not very good at making effective prompts" (#20). These contradictions are perhaps explained by the assignment requirement to consider *both* benefits *and* limitations of ChatGPT for task design purposes; they may also reflect variability in the specific prompt wording used by participants, pointing to a potential limitation if multi-shot prompting approaches are used.[3]

The findings have implications for TBLT in theory and practice. With regard to the task construct, it is clear that teachers need to continue to engage with the definitions of task (such as those proposed by Ellis, 2024) and to apply these definitions to their evaluations of GenAI outputs, rather than taking those outputs at face value. As for the different 'flavours' of TBLT as outlined in Section 2.1, teachers using GenAI should take modes of teaching and learning into account both in their prompt engineering and in their critical evaluation of outputs, in particular if those outputs include suggested pre-task and post-task activities. GenAI therefore has a role to play in task-based pedagogy, but it is a supportive role in partnership with the teachers who remain the ones in control. Essentially, the teacher's role changes from being "producers of tasks" to being "critically literate evaluators of them" (Sun, 2026, p. 423) in the broader context of what are considered, by the teachers, to be contextually-appropriate modes of delivery.

Taking both benefits and limitations into account, it is not surprising that Themes 1 and 6 often coalesced in participant reflections in a 'yes … but' way, making ChatGPT "a real double-edged sword" (#32). Despite the benefits of GenAI, the strong need for teachers to remain in the driving seat and to maintain scrutiny and oversight of GenAI-generated output is reinforced in all the prior studies referred to earlier. It is also emphasised by Sun (2026) who concluded that, ultimately, "the future of TBLT will not be defined by what GenAI can do, but by how teachers imagine and enact its pedagogical possibilities." GenAI may operate as a useful *assistant*, but "only teachers can bring learning to life" (p. 427). The intersection between these two themes provides a useful framework for understanding why AI-human collaboration is necessary within TBLT.

6 Conclusions

Several limitations to the study need to be acknowledged. One limitation is that the assignment brief was to use a zero-shot prompt approach (Kojima et al., 2022) – ChatGPT was told what to create, what theoretical framework to use, what components to include, how long the task phases should take, and what topic and CEFR level to target. No example tasks were provided. This approach was deliberate in

Online First View

that the main purpose of the exercise was not for ChatGPT to design the 'perfect' task, but, rather, for participants to reflect on the benefits and limitations of what ChatGPT could do after a single prompt. It is conceivable that a more iterative process might help generate tasks that are more contextually appropriate and more consistently aligned with TBLT theory. Future research would benefit from evaluating the effects of this kind of process (although, as pointed out in Section 5, this has implications for efficiency). Another limitation is that the study did not differentiate between the participants in terms of, for example, years of teaching experience. From a reflexive TA perspective, such variables are less important than the overall themes constructed from the entire data set. Future research could explore whether specific participant variables (such as years of teaching experience) have a differential impact on perceptions and experiences. If quantification becomes an important component of cross-group analysis, or if greater objectivity is considered crucial to broader generalisation, an analytical approach other than reflexive TA (such as coding reliability TA or codebook TA) would be more appropriate (Braun & Clarke, 2021).

In conclusion, and taking into account that open GenAI use is still in its infancy (Cain, 2024) – ChatGPT was only made publicly available at the end of 2022 – the present study is exploratory. Also, although reflexive TA does not require alignment with previous findings as a form of validation, the findings are, to a large extent, confirmatory, and the data link meaningfully to the wider academic conversation. They add to our emerging knowledge and understanding of how tools like ChatGPT can be used beneficially to support teachers with lesson and resource preparation, as well as limitations that need to be kept in mind.

Importantly, participants in the present study were asked to reflect on their own perceptions of the benefits and limitations of ChatGPT after having used ChatGPT to design a task, without pre-empting what might emerge from that reflection. It is therefore assuring that the perspectives expressed resonate with what has thus far been documented from previous recent studies, thereby helping to substantiate the emerging picture of both possibilities and pitfalls. The present study thus reinforces the *current* message that tools like ChatGPT, whilst very helpful, need to be handled with care (although, no doubt, over time the power of GenAI will increase and its outputs will become more refined). The findings will be helpful for teacher educators who are tasked with guiding teachers as they explore effective practice. In particular, the findings provide encouragement for teacher educators to support the use of GenAI for task generation purposes, including both zero-shot and multi-shot prompting. There must also be opportunities for participants to evaluate and critique the created outputs, whether individually or collaboratively, in line with their own local contexts. This evaluation should draw on the existing literature that defines TBLT and task in theory and illustrates TBLT and task in practice.

Notes

1 Sun's (2026) theoretically-oriented essay provides a very recent exposition of the benefits and limitations of GenAI in the TBLT context; my own personal response to Sun, which draws to some extent on data collected for the present study, has been submitted to *Language Teaching*.

2 I appreciated one reviewer's comment about coding reliability. Following Braun and Clarke's reflexive TA framework, independent coding and inter-coder reliability procedures were not employed. Reflexive TA conceptualises coding as an inherently interpretive and reflexive process, with the researcher actively generating analytic insights through engagement with the dataset rather than uncovering objectively pre-existing meanings within the data. Consequently, coding reliability measures are viewed as philosophically and conceptually inconsistent with reflexive TA. The absence of a second coder and calculations of inter-coder reliability were therefore deliberate methodological choices aligned with the assumptions of the analytic approach.

3 For the assignment, participants were asked to follow a zero-shot prompt approach; however, it was apparent that some participants chose to engage in a more iterative process.

Online First View

Ethics Statement

The study was approved by the University of Auckland Human Participants Ethics Committee, Protocol no. 26125.

References

Belloula, S. (2025). Empowering educators: Leveraging AI to revolutionize lesson planning. *International Journal of Research in Education and Science, 11*(2), 264-280. https://doi.org/10.46328/ijres.1295

Boers, F. (2024). From PPP to APC: Some thoughts triggered by Ellis' (2024) take on task-supported language teaching. *International Journal of TESOL Studies, 6*(4), 34-42. https://doi.org/10.58304/ijts.20240404

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology, 23*(1), 23-41. https://doi.org/10.56297/BKAM1691/WIEO1749

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328-352. https://doi.org/10.1080/14780887.2020.1769238

Bryfonski, L. (2022). Connecting teacher training to task-based language teaching implementation. In M. Ahmadian & M. Long (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 463-477). Cambridge University Press.

Bryfonski, L., & Pineault, C. (2026). The activity design typology: A practitioner-centered approach to understanding task structure. *Language Teaching Research,* 1-31. Advance online publication. https://doi.org/10.1177/13621688261459671

Bui, G. (2024). There is more to the task: A response to Ellis (2024). *International Journal of TESOL Studies, 6*(4), 14-19. https://doi.org/10.58304/ijts.20240402

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends, 68*(1), 47-57. https://doi.org/10.1007/s11528-023-00896-0

Chen, L., & Dan, W. (2024). The supplementary role of TSLT for TBLT: A response to Ellis (2024). *International Journal of TESOL Studies, 6*(4), 62-70. https://doi.org/10.58304/ijts.20240407

Council of Europe. (2001). *Common European Framework of Reference for languages*. Cambridge University Press.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe. www.coe.int/lang-cefr

Dornburg, A., & Davin, K. J. (2025). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL, 37*(3), 332-347. https://doi.org/10.1017/S0958344024000272

East, M. (2022). *Mediating innovation through language teacher education*. Cambridge University Press. https://doi.org/10.1017/9781009127998

East, M. (2024). TBLT from the teachers' perspective: A response to Ellis (2024). *International Journal of TESOL Studies, 6*(4), 20-33. https://doi.org/10.58304/ijts.20240403

Ellis, R. (2015). Teachers evaluating tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 247-270). John Benjamins. https://doi.org/10.1075/tblt.8.09ell

Online First View

Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching, 50*(4), 507-526.

Ellis, R. (2024). Task-based and task-supported language teaching. *International Journal of TESOL Studies, 6*(4), 1-13. https://doi.org/10.58304/ijts.20240401

Erlam, R. (2016). 'I'm still not sure what a task is': Teachers designing language tasks. *Language Teaching Research, 20*(3), 279-299.

Gurzynski-Weiss, L., & IATBLT. (n.d.). *The TBLT language learning task bank*. https://tblt.indiana.edu

Han, Z. (2018). Task-based learning in task-based teaching: Training teachers of Chinese as a Foreign Language. *Annual Review of Applied Linguistics, 38*, 162-186. https://doi:10.1017/S026719051800003X

Hasnain, S., & Halder, S. (2023). Exploring the impediments for successful implementation of the task-based language teaching approach: A review of studies on teachers' perceptions. *The Language Learning Journal, 51*(2), 208-222. https://doi.org/10.1080/09571736.2021.1989015

Kerr, R. C., & Kim, H. (2025). From prompts to plans: A case study of pre-service EFL teachers' use of generative AI for lesson planning. *English Teaching, 80*(1), 95-118. https://doi.org/10.15858/engtea.80.1.202503.95

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *arXiv*, 1-42. https://doi.org/10.48550/ARXIV.2205.11916

Long, M. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences, 379*(1), 259-278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Long, M. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics, 36*, 5-33.

Lu, J., Ma, Q., & Gao, X. (2026). Adapting TBLT in Chinese secondary schools. *ELT Journal,* Advance online publication, 1-14. https://doi.org/10.1093/elt/ccag029

Moorhouse, B. L. (2024). Beginning and first-year language teachers' readiness for the generative AI age. *Computers and Education: Artificial Intelligence, 6: 100201*, 1-8. https://doi.org/10.1016/j.caeai.2024.100201

Moorhouse, B. L., Ho, T. Y., Wu, C., & Wan, Y. (2025). Pre-service language teachers' task-specific large language model prompting practices. *RELC Journal*, Advance online publication, 1-22. https://doi.org/10.1177/00336882251313701

Pham, N. T., & Nguyen, H. B. (2018). Teachers' perceptions about task-based language teaching and its implementation. *European Journal of Foreign Language Teaching, 3*(2), 68-86. https://doi.org/10.5281/zenodo.1241474

Sun, Q. (2026). Rethinking the role of teachers in generative AI-supported task-based language teaching. *Language Teaching*, *59*(4), 422-428. https://doi.org/10.1017/S0261444826101372

Teng, M. F. (2024). "Can we play with tasks?": A response to Ellis (2024). *International Journal of TESOL Studies, 6*(4), 43-54. https://doi.org/10.58304/ijts.20240405

Van den Branden, K. (Ed.). (2006). *Task-based language education: From theory to practice*. Cambridge University Press.

Van den Branden, K. (2016). The role of teachers in task-based language education. *Annual Review of Applied Linguistics, 36*, 164-181. https://doi:10.1017/S0267190515000070

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Online First View

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.

Willis, J. (1996). *A framework for task-based learning*. Longman Pearson Education.

Zhu, Y. (2024). Designing tasks in ELT textbooks for young learners: A response to Ellis (2024). *International Journal of TESOL Studies, 6*(4), 71-74. https://doi.org/10.58304/ijts.20240408

Martin East is Professor of Language Education in the School of Education and Social Practice, the University of Auckland, New Zealand. He has published widely in the fields of language pedagogy and language assessment. His books include two works focused on TBLT: *Task-based language teaching from the teachers' perspective* (John Benjamins, 2012) which examined how teachers understand and enact task-based ideas, and *Foundational principles of task-based language teaching* (Routledge, 2021) which presents an accessible and reader-friendly exploration of the core theories and central practices of TBLT. He served for five years as President of the International Association for Task-Based Language Teaching (IATBLT).

Online First View