

Negotiating Educational Authority in Human-AI Collaboration: Exploring Chinese Language Teachers' Agency in AI-Assisted Task Design

Xinzhu Sun

International College, Macau University of Science and Technology, Macau 999078, China

Abstract: Generative AI (GenAI) can be used to create lessons and tasks, but its pedagogical effectiveness depends on teachers' professional judgments about whether to accept, modify, or reject AI suggestions. This qualitative multiple-case study investigates the professional agency of six Chinese language teachers who created or revised beginner-level shopping tasks with ChatGPT. Data comprised semi-structured interviews, time-stamped ChatGPT interactions, and teacher revisions, allowing comparison between stated beliefs and observable design decisions. Reflexive thematic analysis and cross-case comparison identified three interrelated patterns. First, teachers filtered AI recommendations through existing pedagogical knowledge and beliefs, including judgments about learner level, syllabus content, instructional sequencing, and task-based language teaching (TBLT). Second, teachers localized AI-generated tasks to learners' interests and classroom ecology, especially by reducing cultural distance and simplifying classroom-management demands. Third, efficiency shaped uptake: some teachers were highly selective and retained little AI-generated content, whereas others prioritized efficiency and accepted most of it. The findings conceptualize AI-assisted task design as a negotiation of pedagogical authority rather than simple technology adoption. GenAI expanded teachers' options and reduced preparation demands, but teachers still had to ensure contextual, theoretical, linguistic, and practical fit. The study contributes process-level evidence to emerging research on language-teacher agency and GenAI and to technology-mediated TBLT by showing how task-design principles are applied, adapted, or neglected in human-AI interaction.

Keywords: Generative AI, ChatGPT, Teacher Agency, Task-based Language Teaching, Chinese Language Teaching, Instructional Decision-Making, Task Design.

1. Introduction

Generative artificial intelligence has developed rapidly from an emerging technology into an everyday tool for many language teachers. Current systems can generate lesson plans, examples, role-plays, information-gap tasks, feedback, and differentiated content within seconds. Since 2023, research has increasingly shifted from asking whether language educators will use GenAI to examining how they use it. Recent studies and reviews report both perceived benefits, including speed, ideation, personalization, and resource generation, and concerns about accuracy, overdependence, pedagogical suitability, ethics, and professional competence [1–4]. Intervention research further suggests that explicit teacher education can strengthen professional GenAI competence, particularly pedagogical competence and critical awareness [5].

This emerging literature is valuable, but much of it relies on self-report or perceptions of intended practice. What teachers say they value may not fully predict what they do when an AI system proposes a concrete instructional action. That moment is especially important in task design. A model may suggest a role-play, add vocabulary, change the order of input and output, introduce competition, or propose a task sequence. The teacher must then decide whether the suggestion is appropriate for particular learners, goals, time constraints, curricular expectations, and theories of language learning. This is not simply a question of technology acceptance; it is an exercise of professional agency in practice.

The issue is especially relevant to task-based language teaching (TBLT). TBLT treats communicative tasks as central units of planning and emphasizes meaning-oriented activity,

communicative outcomes, sequencing, needs, and principled attention to form [6–8]. Research on technology-mediated TBLT has long emphasized that technology should be integrated with pedagogical principles rather than attached superficially to tasks [9]. GenAI makes this relationship more complex because the technology can now propose the task itself. The question is therefore not only what the technology affords, but how teachers negotiate pedagogical authority with a system that can generate plausible instructional designs.

The current study addresses this question through six Chinese language teachers who created or revised beginner-level shopping tasks with ChatGPT. It combines semi-structured interview data with interaction logs and post-generation revision records. This design makes it possible to examine teachers' actual uptake, rejection, localization, or transformation of AI-generated content alongside the rationales they provide. The study asks three questions: (1) How do Chinese language teachers evaluate task suggestions produced by GenAI? (2) How are these evaluative criteria enacted in visible revisions of AI-generated tasks and materials? (3) What do differences in uptake reveal about teacher agency in AI-assisted task design?

2. Literature Review and Conceptual Framework

2.1. GenAI in Language Education: From Affordances to Professional Judgment

The systematic literature review is a fast growth of the GenAI studies in language learning and teaching. The recent reviews have pointed out to the advantages of content creation, feedback, personalization, interaction, and support of learners

but also to the threats of reliability, privacy, bias, integrity, and low-quality pedagogy [1][10]. Teacher-focused research is particularly relevant because it shows that access to GenAI does not in itself ensure effective use; teachers require sufficient professional knowledge and critical judgment to evaluate and adapt AI output [2–5,11].

Professional knowledge is being more and more the subject of study than access to it, which is the determining factor in teacher-focused research. The readiness of a new language teacher was not equal (Moorhouse [2]), as well as the readiness of the first-year teachers were not equal either (Moorhouse and Kohnke [3]). In their article, Ma et al. [4] have developed a six-construct framework that conceptualized ChatGPT literacy in teaching languages and then tested this concept with 492 language teachers in 41 countries. Moorhouse et al. [5] later demonstrated how training could be used in order to achieve professional competence in GenAI and even demonstrate a significant level of competence in pedagogical skills and critical thinking.

The more recent work tends to the situated decisions, which are key to the current research. The systematic review by Eedelouei directly conceptualizes the judgments of language teachers regarding GenAI in terms of teacher agency and defines practical-evaluative decision-making as a fundamental yet underexplored process [12]. Studies on teacherness also support that contextual and pedagogical judgment cannot be dispensed with even when GenAI produces materials effectively [13]. These trends indicate that it is not general acceptance of AI that is the most informative unit of analysis but rather the order of judgments whereby teachers make their choices on what to retain, modify or dismiss.

2.2. Teacher Cognition and Ecological Agency

Studies of teacher cognition have been a long time in demonstrating that classroom action is mediated by what teachers know, believe and think, such as previous learning experiences, teacher education and accumulated classroom practice [14]. This view assists to understand why the same AI recommendation can be perceived differently among teachers who have different instructional backgrounds. A technically plausible output has no definite pedagogical worth; it is assessed based on the existing repertoire of beliefs and situated knowledge of a teacher.

This is a cognitive account with the agency that has been added to it. Priestley, Biesta and Robinson are not the proponents of the concept of teacher agency as something inherent in individuals but rather the outcome of the interaction between history and future and present circumstances [15]. In AI-assisted design, GenAI is part of such an ecology: it makes it easier to choose from more options, reduces costs of production, and it can also create limitations (inadequate levels of examples, or too much pedagogical effort). Teachers are able to see the affordances and restrictions on what they do with them, and this is how we can observe the agen.

2.3. TBLT as a Pedagogical Filter for AI-Generated Designs

TBLT is a rather useful perspective as it makes the principles of design visible. The tasks will usually be done in such a way that learners are required to employ language in terms of meaning and the purpose, and to accomplish some result; whereas, when it comes to implementation, there will be an analysis of needs, task complexity, sequence, interactional factors, and form [6][7] [8] [16]. TBLT technology-mediated approach takes this concept further into the digital world and proposes that technologies can be used in pedagogically significant ways with the task design [9]. Meta-analytical studies have shown that GenAI-based TBLT may be effective in generating positive learning experiences and conditions of task planning and AI design moderate the results [17].

In the case of the current study, TBLT is not considered as a checklist that all participants were supposed to adhere to. Rather, it serves as one of several possible professional filters. One of the participants specifically assessed AI recommendations based on TBLT principles like task complexity, sequencing and focus on form; other participants focused on syllabus alignment, learner engagement, local cultural fit, established sequencing beliefs or preparation efficiency. The comparison of these filters shows how pedagogical authority is distributed differently in the cases.

3. Method

3.1. Design and Participants

Table 1. Case profiles derived from interview and interaction records

Case	Salient context	AI orientation	Primary evaluative filter	Observed uptake pattern
T01	Experienced beginner-level teacher; 10+ years	Occasional, preparation-focused	Learner level + syllabus progression	High filtering; deletes or simplifies beyond-progress items
T02	Adult small-group teaching context	Frequent; flexible in-class use	Engagement + willingness to speak	Experimental uptake; tries and retains engaging designs
T03	20+ years' experience	Skeptical	Accuracy + input-before-output belief	High filtering; accepts lexical/example support, resists structural change
T04	Secondary-school teaching context	Idea-generation tool	Local cultural fit + manageability	Selective localization; replaces distant scenarios/content
T05	Doctoral student; strong theoretical orientation	Efficiency with theory-led evaluation	TBLT: complexity, sequencing, focus on form	Theory-mediated revision; requests explicit restructuring
T06	Part-time teacher; limited planning time	Strongly positive	Efficiency + basic acceptability	Very high uptake; minimal modification

The study used a qualitative multiple-case research design which included six teachers of the Chinese language who were identified as T01, T02, T03, T04, T05 and T06. The cases are to be picked in such a way that there is diversity in the teaching experience, context of institutions, theoretical approach, and knowledge about GenAI. In this case, an

experienced beginner teacher was T01 with over ten years of teaching career, T03 with more than two decades of teaching career and had a high level of doubt in AI, T05 with a doctorate in education, and explicitly referring to TBLT theory, and T06 with part-time teaching experience, and a lot of preparation time. To avoid adding unsupported

demographic detail, T02 is described in terms of an adult small-group teaching context and frequent, flexible AI use, while T04 is described in terms of a secondary-school context and strong attention to local cultural relevance and classroom manageability. Unnecessary demographic detail is omitted to protect anonymity.

3.2. Data Sources and Procedure

There were two sources of data. In the first case, the participants were subjected to a semi-structured video interview on-line and it took about 55 minutes. The interviews were recorded in audio and written down as they happened. It was the transcription practice that allowed to keep pauses, fillers, laughter, discourse markers; certain parts of Chinese were translated into English with their pragmatic power kept. The analysis excerpts were based on the use of AI, criteria used to assess and the decision-making process of task-revision.

Second, participants engaged in standardised ChatGPT interactions within a ChatGPT Plus web environment. The interface read GPT-5. Records cover March 15th-April 5th 2026 and retain prompt-response order, session identifiers, and subsequent outputs presented to participants. Post-generation edits are recorded in the teacher-retained or teacher-submitted versions of the task found in the dataset. To render behavioural evidence auditable without assuming an unrecorded capture technique, each relevant AI output was compared to the corresponding teacher-edited version and observable operations (e.g., retention, deletion, lexical simplification, contextual substitution, structural revision, elaboration) logged. Thus, the analysis does not infer latent cognitive processes from the model log alone, but relates what teachers said in interview to changes apparent between the AI output and the recorded revised task.

All the participants were assigned a similar shopping-task domain so that all of them would be able to interpret the differences between the cases in terms of a similar design issue. Simultaneously, there was enough freedom in the prompts to enable teachers to accept, refuse, shift or extend suggestions of the model. The resulting traces can then be considered as the design interaction and not the performance of models.

3.3. Analytic Approach

It was analyzed in the form of a reflexive thematic analysis [18][19] and cross-case comparison. The coding was done in four stages. The first step is to code the interview data into rationales (e.g., the level, curriculum, theoretical justification, involvement, cultural understanding, accuracy, efficiency). Second, interactional information were coded on teacher activities (retain, delete, lexical simplify, contextual substitute, structural revision, elaborate, accept with no change). Third, the rationale behind the interviews and actions of interactions were connected in the cases where there was convergence or divergence between what they said and their behavior. Fourth, the patterns were compared in the cases to produce more general themes about filtering, contextualization and efficiency of uptake.

It was not aimed at computing inter-rater reliability, which would be incongruous with reflexive thematic analysis as interpretive knowledge production [19]. Analytic transparency was enhanced instead by preserving an evidence chain of transcript excerpt to interaction trace to analytic claim. The claims are intentionally limited to the interview

and interaction evidence that is presented in the dataset.

3.4. Ethics and Data Integrity

The identities of the participants were kept anonymous (T01-T06), and in the records of the interactions, there was no personal information. The subjects gave a verbal consent to be involved in the study and to use anonymized interview as well as interaction data in the research report. Every piece of writing and behavior that has been reported in the manuscript was treated anonymously. In the analysis, it is possible to separate evidence created by the participants and interpretation made by researchers, which does not imply any intentions or beliefs of the AI system.

4. Findings

Three interlocking themes captured the negotiation of pedagogical authority across the six cases. The themes were not mutually exclusive categories of teachers; rather, they represented recurring dimensions of decision-making that emerged with varying intensity across cases. Table 2 summarizes the main cross-case pattern before the themes are developed in detail.

Table 2. Cross-case pattern of AI uptake and pedagogical filtering

Case	Dominant filter	Typical action toward AI output	Relative uptake
T01	Level and syllabus	Delete/simplify beyond-progress language	Low-moderate
T02	Engagement	Try, retain, and adapt interactional ideas	Moderate-high
T03	Accuracy and sequencing belief	Retain examples; reject structural reordering	Low-moderate
T04	Local fit and manageability	Substitute contexts; simplify classroom demands	Moderate
T05	Explicit TBLT principles	Restructure prompts and task sequence	Moderate
T06	Efficiency	Accept if basically usable; minimal edits	Very high

4.1. Theme 1: Teachers Filtered AI through Existing Pedagogical Beliefs and Knowledge

The AI output was not often considered as inherently authoritative in the cases. It was assessed by participants based on knowledge gained through experience of teaching, curriculum pacing, learning beliefs or formal pedagogical theory. The most powerful filtering was observed in T01, T03 and T05, however, the nature of the filtering was quite different.

4.1.1. Learner Level and Syllabus Progression as Non-negotiable Constraints

“I definitely look first at whether the difficulty is appropriate. I’ve taught beginner classes for more than ten years, so I know what students at this level can handle. AI sometimes gives you things that look lively, but the vocabulary is beyond the level and the sentence patterns are too complex. Students simply cannot digest them.” (T01) [Translated from Chinese]

T01’s interview account was directly mirrored in the interaction record. Asked to generate five HSK2-level

shopping gap-fill items without out-of-syllabus vocabulary, ChatGPT produced items that included 贵 ('expensive') and 试 ('try on'). T01 deleted those items because they had not yet been taught and replaced them with simpler material. The important point is not whether those words are globally appropriate to an HSK2 label; T01's criterion was the locally taught syllabus and current instructional sequence. Her agency was therefore enacted as curricular gatekeeping: the model's generic proficiency estimate was subordinated to knowledge of what this particular class had encountered.

"It also has to fit the teaching syllabus. We teach according to a schedule and a plan. If its suggestion goes beyond where we are, then of course I can't use it." (T01) [Translated from Chinese]

4.1.2. Experience-based Beliefs Could Enable Principled Resistance

Learning language certainly has input before output. Its design required students to speak first and explain the points of language later. That is not according to the learning process. (T03) [Translated from Chinese]

T03 was not so much opposed to it, because he had a theory of practice that was well entrenched in the last 20 years. T03 did not challenge the pace of generation of examples as shown by the trace of ChatGPT but half of the eight shopping expressions were eliminated and one of the proposed expressions (I'll buy this one) was replaced with a simpler one (I want this). In the model suggestion to have an output-first sequence of activities, T03 would not agree to such a structural shift. The example demonstrates that the agency can be expressed through resistance in cases where the teacher feels that the AI is suggesting the wrong way of teaching the process of learning.

Simultaneously, T03 did not say no to AI. The instances of the words that were considered to be accurate and on a level were left as they were. It was practical: it was accepted as an instant producer of possible examples, but it was not given power over the sequence of instruction. This is the kind of division of labor in which teachers are either adopters or resisters.

4.1.3. Formal Theory Could Be Used to Interrogate and Redirect AI

T05 displayed a different form of high filtering. Rather than relying mainly on accumulated classroom routines, T05 explicitly invoked TBLT concepts. In the interaction record, T05 asked the model to revise a shopping task so that the communicative outcome was clearer, the progression in cognitive and linguistic demand was more controlled, and a brief post-task focus-on-form stage was added. T05 did not simply edit surface wording; the prompts required the model to restructure the task around an explicit pedagogical rationale.

"AI gives me a draft very quickly, but I still need to ask: what makes this a task, where is the outcome, what is the complexity, and when should we draw attention to form? Otherwise it can look communicative without being well designed." (T05) [Translated from Chinese]

In this case, the theoretical knowledge was used as a means of interrogation of AI results. The model had increased the number of designs and T05 still had the control over the criteria of tasks. This trend is consistent with the overall argument that technology must be made to take second place to task principles and not be pedagogically self-justifying [9][16].

4.2. Theme 2: Teachers Recontextualized AI for Learner Interests and Classroom Ecologies

Another cross-case pattern was contextual fit. The participants did not assess AI recommendations based on linguistic accuracy or theoretical coherence; they also thought about whether the situation, objects, pacing and interactional demands would be compatible with certain learners in certain classrooms. This ecological dimension was made particularly visible by T02 and T04.

4.2.1. Cultural Familiarity Mattered for Perceived Authenticity

T04 substituted the examples that were usable and yet were not in the culture of the learners. The shopping was retained as a communicative activity but the goods, places and situations to be used in the shopping were changed into those which would be more familiar to the secondary-school students in the given environment. Such changes were not superficial. It changed the level of background explanation needed and it also made the learners more likely to have common knowledge when interacting with each other.

This trend indicates that authenticity was considered as ecological instead of universal. A scenario created by an AI can be similar to a realistic task in the real world and yet appear to be unauthentic to a certain class. The revisions made by T04 thus moved pedagogical authority out of the generalized cultural script of the model into the knowledge of local learners possessed by the teacher. The model provided a transferable skeleton of tasks; the teacher determined what constituted a meaningful social situation.

4.2.2. Engagement Was Treated as a Design Criterion, Not a Decorative Add-on

T02's adult small-group context produced a different form of contextualization. T02 showed comparatively high willingness to trial AI-generated interactional ideas when they appeared likely to increase willingness to speak. Rather than asking only whether an activity was theoretically ideal in the abstract, T02 evaluated whether it could create enough immediacy, choice, or interpersonal relevance to sustain participation among adult learners. Ideas judged potentially engaging were more often tried or retained, whereas elements that risked becoming overly scripted were adapted.

The important analytic point is that T02's relatively open stance toward AI was still conditional. Uptake was not evidence of absent agency; it reflected a different pedagogical priority. Whereas T01's threshold question was 'Has this language been taught?', T02's was closer to 'Will this make these learners talk?' Different filters produced different patterns of acceptance from the same broad technological affordance.

4.2.3. Classroom Manageability Constrained Otherwise Attractive Designs

T04 also put some filters on the suggestions, based on what would be needed in the classroom. For example, the AI may have multiplied roles and phases or resources to be used. T04 is a simplification of the designs that had less moving parts but the same communicative exchange. In this situation, the teacher was not only aware of the language of the learners, but he or she was also conscious about time arrangement, grouping, instructions and transitions.

These changes show that the evaluation of AI-generated tasks at the level of written plans is not sufficient. A plan can

seem rich on the page but too much management load in practice. The teacher agency was seen when an abstract task design was translated into a classroom event that could be implemented. This translation work is important to ecological accounts of agency because what is possible depends on local material and organizational conditions [15].

4.3. Theme 3: Efficiency shaped uptake without eliminating teacher agency

Efficiency was present in all six cases, although it influenced uptake most strongly for T06. Overall, participants appreciated GenAIs potential to decrease the amount of time needed to produce examples and alternative ideas for tasks. The key distinction was how much pedagogical examination they were prepared to, or capable of, undertaking post-generation.

4.3.1. Efficiency Could Function as a Constraint on Depth of Evaluation

T06, a teacher who is not full-time and has limited time to plan was the one that received more than half of the AI output as compared to other teachers. It had been found that once it was determined that the suggestions were at least in principle applicable, there was little revision involved in the interaction trace. This did not imply that T06 lacked standards but on the contrary the level of acceptance was relatively large and it was greatly influenced by the practical aspect of completing preparation within the shortest possible time. There was also pedagogical power, only in the form of a less demanding screening system.

This case is significant as it demonstrates the teacher agency does not imply the ability to edit a lot. The time-efficient teacher would be able to take an agentic choice of leaving much of the design work on AI in order to save time. Meanwhile, there is a possible weakness that can be seen in the case: when the cost of verifying increases and the cost of accepting decreases, then mistakes or mismatches or theoretically weak systems may be allowed to go through without being examined too closely. This was because efficiency was both a condition and an affordance which was redistributing the labor of evaluation.

4.3.2. High Uptake and High Agency Were Not Opposites

T02 was also relatively high in uptake but with other reasons. The frequent use of AI and readiness to experiment by T02 implied that the model recommendations could be regarded as temporary classroom hypotheses: when an idea seemed interesting and viable, it was worth attempting. In this case, agency was evident in experimental choice instead of rigorous pre-implementation screening. On the contrary, T01 and T03 acted through rejection and simplification whereas T05 acted through theory-based re-prompting. The cross-case comparison thus suggests that the visible editing is not a perfect proxy of agency.

4.4. Cross-case Synthesis: Negotiated Pedagogical Authority

All in all, the cases indicate that there is a spectrum of pedagogical filtering and not just one form of AI. In the case of T01 and T03, the model was placed as an inferior tool which had to be verified over and over again in terms of its recommendations on syllabus advancement, level, correctness, and the existing sequencing assumptions. The power of T05 remained very high but it was also based on the use of prompting the theory-based approach where the model

was brought into a more direct dialogue with the design process. T04 localized the output of AI to the cultural and organizational ecology of the classroom. T02 employed the experimental rule of engagement. T06 put the biggest part of the designing job since saving time was one of the most important professional requirements.

It is possible to identify three features that are common in the concept of negotiated pedagogical authority, which can be seen throughout this change. To begin with, GenAI did not become a tool of design as an impartial object: it was able to provide ready and full recommendations on what could be done next, and hence the options were already presented to it. Second, the teachers were not passive in adopting these choices; they made their own assessments of them through different practical and theoretical lenses. Third, there was dynamicity in terms of distribution of power among the stages of work. The AI usually took over during the first phase of ideation since it was capable of coming up with alternatives at a very fast rate whereas teachers had the ability to make decisions when needed to consider context, curriculum, theory or feasibility. The outcome of such an arrangement was neither man-dominated in the sense of human nor AI-managed, but rather a compromise between the two sides of design labor.

5. Discussion

This paper investigated the evaluation and revision of GenAI-generated beginner shopping tasks by six Chinese language teachers. The research questions are answered by three findings. Teachers applied heterogeneous criteria - level of learners, progression of syllabus, sequencing beliefs, TBLT principles, engagement, cultural fit, manageability, accuracy, and efficiency to rate AI suggestions. These standards could be seen in actual revision actions such as deletion, simplification, contextual substitution, restructuring, and selective retention. The differences in uptake did not correspond well with greater or lesser agency. Rather, delegation was manifested in various forms of agency such as resistance, experimentation, and redirection.

5.1. From Technology Acceptance to Practical-evaluative Judgment

The first contribution is conceptual. The attitude, the willingness to use, the literacy and readiness have been discussed in a number of early discussions of GenAI as an education [2][3][4]. It is important that these constructs are still relevant, but the current results demonstrate why they do not suffice when it comes to design practice. A teacher may be optimistic about GenAI and yet disapprove of a specific proposal; a sceptical teacher can learn useful words in the process; a frequent user can be very sensitive to the context of his or her usage; a theory-based user may make prompting an act of teaching. It is not the general attitude towards technology but a local position on the issue of whether the output is compatible with a particular ecology of instruction.

This view of the process is quite close to the argument presented by Eedelouei [12] that language teacher agency in GenAI situations is based on practical-evaluative judgment. The current paper develops this argument further, demonstrating how such a judgment can be seen in revision traces. Agency may be researched not only based on what teachers report but also on material differences between model suggestions and designs retained by teachers. This is

methodologically important since it minimizes reliance on retrospective self-report and provides the research on teacher-agency with a more detailed behavioral unit of analysis.

5.2. Negotiated Pedagogical Authority as a Form of Human-AI Division of Labor

The second is the idea of negotiated pedagogical power. It is not only that the GenAI systems will be able to make familiar tasks faster, but they can also recommend a system of teaching and learning, an order of presentation, the way of communication, as well as content. This implies that the design authority is in part externalized. The cases however indicate that it was not complete and permanent in its nature. There was no total and stable externalization: when local knowledge was needed, teachers were again given authority to know what had already been taught by learners, what would be meaningful in terms of cultural references, how much coordination between classes could be achieved, or even whether a task was good enough to be performed according to some pedagogical theory.

This trend echoes the recent literature on teacherness, which contends that GenAI can produce content quickly but does not possess contextual pedagogical judgment of professional teaching [13]. The current research contributes a complementary perspective: contextual judgment is not simply a leftover human ability that is left over after automation. It actively decides what parts of AI output become instruction. Teacher expertise thus operates as a selection and transformation process whereby generalized model output is transformed into locally legitimate pedagogy.

The idea also explains why high uptake should not necessarily be viewed as deskilling. The massive delegation of T06 can be interpreted as an agentic reaction to time scarcity whereas the relatively open uptake of T02 is a manifestation of experimental orientation. Nevertheless, negotiated authority may be unbalanced when the practical price of scrutiny is excessive. In that respect, workload conditions are not peripheral to responsible AI use; they determine how much evaluative work teachers can realistically do. GenAI claims that it saves time, but attention must be paid to the time needed to verify, localize and check pedagogically.

5.3. Extending Technology-mediated TBLT: from Digital Affordances to AI-generated Task Proposals

The third contribution is the TBLT. The principles of alignment between the technological features and task design have been put in focus by technology mediated TBLT [9]. GenAI makes it different as the technology will be able to come up with a proposal of the task. It raises the necessity of differentiating surface communicativeness against the coherence of tasks-theoretical. The case of T05 was that one activity could appear to be interactive, but still need to be modified to make the outcome more explicit, to adjust the complexity or to shift the attention to something more deliberate.

The results thus generalize technology-mediated TBLT as a question of how to apply tasks with the help of technology into a question of how teachers assess task proposals generated by technology. This difference is important since GenAI can shorten ideation and drafting to seconds, so that poor design choices look polished and ready to use. The

principles of TBLT may be used as diagnostic standards that will slow this process down at key points. Meanwhile, the other cases warn against the assumption that explicit knowledge of TBLT is the only valid basis of judgment. Knowledge of syllabus, engagement of learners, cultural familiarity, and manageability in the classroom were also equally influential in deciding whether a task could work or not.

The recent study of Li, Shadiev and Chiu [17] has revealed that TBLT can be effective in learning as a result of the GenAI-based technology and the effects of the instructional design are moderate. The existing qualitative literature supports such an outcome-oriented research as well since it demonstrates who does the mediation of the design: teachers, with the same choices of complexity, sequencing, local relevance, feasibility etc. TBLT based on GenAI must then be considered to be not only an AI implementation of a task-based lesson but also a teacher-mediated design process.

5.4. Implications for Teacher Education and Professional Development

The implications of the results are immediate to the education of teachers. First, it is necessary that the GenAI competence involves the analysis of the output in the pedagogy level as well as not just in the prompt building or hallucination consciousness. In the six-construct ChatGPT literacy model suggested by Ma et al. [4], there is an evaluation and ethics element; the current examples imply that the evaluation should be carried out with a real teaching object and the local curriculum limitations. The teacher must have chances to make comparisons between possible AI generated work, detect assumptions and then decide on the retain-revise-reject decision.

Second, teachers should be subjected to a variety of professional filters. The level and syllabus are not the only ones that need to be considered but also the outcome of the task, the complexity of the language and thought process, the sequence of the tasks, the cultural aspect of the work, the accessibility of the task, the cost of verification, etc. The intervention data has already proved that the explicit training can enhance the competence of GenAI in professional life [5]. It would be helpful to have design studios where the teacher is given an AI-generated assignment, which he/she edits to suit different learners and then explains why each version will vary.

Third, the issue of delegation as a professional choice should be discussed in teacher education. It is not aimed at the maximization of editing or reduction of AI use. Instead, teachers are to be taught how to make choices on which design functions they can safely delegate and which ones need human supervision. As an example, idea generation might be low risk, whereas syllabus alignment, linguistic accuracy, assessment consequences, or culturally sensitive content might need more attention. This division of labor may be made explicit in order to avoid uncritical acceptance and blanket rejection.

5.5. Implications for the Design of GenAI Tools for Language Teachers

The results also indicate design trends of AI systems. Existing chat interfaces tend to be optimized to complete smoothly as opposed to being pedagogically traceable. Agency might be more supported by teacher-facing tools that enable the user to define the taught syllabus, learner profile,

forbidden vocabulary, lesson time, local context, and task-design principles as ongoing constraints. Systems might also render uncertainty more apparent and offer options with reasons instead of a single refined response.

No less significant, the interfaces might be helpful in the process of comparison and not replacement. A teacher could ask for three options of a task to be done in order to reduce the linguistic burden on the student, increase the information gap between them or make it easier to manage in the classroom, and then check the trade-offs. The design space generator will be GenAI with such features and at the same time, the role of the teacher as an adjudicator in the pedagogical field is preserved. It would fit better than the pattern of automation that is being observed here where the most efficient lesson plan is created in a single move.

6. Limitations and Future Research

The research had a number of restrictions. It was based on the sample of six teachers and one task domain. The multiple-case study is not a statistical generalization but an analytic comparison. Shopping tasks are good in that they can be used to make similar design problems of communication, but other aspects such as writing, pronunciation, assessment, project work or even more advanced content teaching can produce various types of teacher-AI bargaining.

Second, the data are not the actual interview stories but rather the ChatGPT records and notes on the changes made in the classroom. A teacher can plan a task of one type and change it once more when teaching to adapt to the behavior of learners. In the future research, the traces of design should be related to video recording of the classroom, stimulated recall, interaction between students and after-class reflection so as to trace agency within planning and implementation.

Third, the cases differ in experience, setting, theoretical orientation, and time pressure but the study was not set up to isolate the causal effects of any one factor. Larger comparative studies might explore whether professional experience, institutional workload, AI literacy, or formal TBLT knowledge systematically predicts certain patterns of delegation and filtering.

Fourth, GenAI systems are evolving quickly. The interactions discussed in this section took place on a ChatGPT Plus interface with GPT-5 being shown in March-April 2026. It is possible that model behavior, interface affordances and institutional expectations may vary, i.e., certain output tendencies cannot be considered as fixed characteristics of GenAI in general. The longer-lasting impact is the analytic emphasis on teacher judgment and negotiated authority.

The current qualitative classifications can also be operationalized in the future to conduct a mixed research. An illustration of this is that revision traces would have been coded at large, with retention, deletion, substitution and restructuring, which are then associated with the teacher beliefs or task-quality judgments. The above work will aid in establishing whether it is possible to find the continuum seen here in the language, education field as well as model ecosystems.

7. Conclusion

GenAI is capable of producing language tasks at a pace that was not previously possible, yet speed does not eliminate the necessity of pedagogy. In six Chinese language teachers, AI-generated suggestions could only be used after making

judgments regarding the level of learners, syllabus progression, sequencing, theoretical coherence, engagement, cultural relevance, manageability, accuracy and time. Such decisions were apparent in what the teachers kept, deleted, simplified, localized, restructured or accepted with little alteration.

The study thus suggests that AI-assisted task design is most appropriately conceptualized as negotiated pedagogical authority. GenAI increases the option space and may assume significant drafting work, but it does not decide what is pedagogically appropriate in a specific classroom. That legitimacy is granted by teachers in their professional judgment in situ. Notably, agency can be articulated via rejection, theory-driven redirection, contextual substitution, experimental uptake or deliberate delegation under time pressure. The key question is not how much of the AI output survives, but how and why authority is distributed throughout the design process.

In the case of language teacher education, the implication is obvious: responsible use of GenAI will not be achieved by prompt fluency. The teachers require training in assessing refined products against curriculum, learners, principles of tasks and classroom realities. In the case of TBLT, the appearance of AI-generated task proposals does not make teacher mediation less, but more significant. The future of AI-based language instruction will not rely on whether or not GenAI can create plausible tasks, but whether teachers are enabled to use informed judgment over the tasks that reach students.

Declarations

Ethics and consent: The participants were anonymized as T01-T06 and gave verbal informed consent to participate and use the anonymized interview and interaction data in research reporting.

Availability of data: Due to the fact that the dataset has interview transcripts and interaction records of a small identifiable professional sample, access should be controlled by consent conditions and anonymization requirements that are relevant to the study.

References

- [1] Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34(1), 335-359. <https://doi.org/10.1080/10494820.2025.2498537>
- [2] Moorhouse, B. L. (2024). Beginning and first-year language teachers' readiness for the generative AI age. *Computers and Education: Artificial Intelligence*, 6, 100201. <https://doi.org/10.1016/j.caeai.2024.100201>
- [3] Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System*, 122, 103290. <https://doi.org/10.1016/j.system.2024.103290>
- [4] Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. *Computers and Education: Artificial Intelligence*, 7, 100278. <https://doi.org/10.1016/j.caeai.2024.100278>
- [5] Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L. M. A., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, 125, 103399. <https://doi.org/10.1016/j.system.2024.103399>

- [6] Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- [7] Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.
- [8] Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- [9] Gonzalez-Lloret, M., & Ortega, L. (Eds.). (2014). *Technology-mediated TBLT: Researching technology and tasks*. John Benjamins. <https://doi.org/10.1075/tblt.6>
- [10] Al-khresheh, M. H. (2024). The future of artificial intelligence in English language teaching: Pros and cons of ChatGPT implementation through a systematic review. *Language Teaching Research Quarterly*, 43, 54-80. <https://doi.org/10.32038/ltrq.2024.43.04>
- [11] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, 5, 100156. <https://doi.org/10.1016/j.caeai.2023.100156>
- [12] Eedelouei, S. (2026). Factors influencing language teachers' judgements and decision-making about the use of generative AI tools: A systematic review. *Teaching and Teacher Education*, 171, 105333. <https://doi.org/10.1016/j.tate.2025.105333>
- [13] Chen, Q., Tawilapakul, U., & Lin, M. Y. A. (2026). Teacherness in the age of GenAI: An ethnographic exploration of pedagogical decision-making and AI integration in English language teaching. *System*, 136, 103877. <https://doi.org/10.1016/j.system.2025.103877>
- [14] Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81-109. <https://doi.org/10.1017/S0261444803001903>
- [15] Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury Academic.
- [16] Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781108643689>
- [17] Li, Y., Shadiev, R., & Chiu, T. K. F. (2026). Applying generative artificial intelligence to task-based language teaching and learning: A systematic review and meta-analysis. *TechTrends*, 70, 150-163. <https://doi.org/10.1007/s11528-025-01140-7>
- [18] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- [19] Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.