USING AI APPS TO ASSIST TEACHERS IN DESIGNING READING OR LISTENING TASKS FOR GIFTED STUDENTS

TABLE OF CONTENT

Chapter 1. Introduction

1. Context and Rationale:

Gifted learners, particularly those engaged in advanced language programs, require exposure to cognitively demanding, authentic, and stimulating input to sustain their intellectual curiosity and promote higher-order thinking skills. In the domains of reading and listening, this necessitates the use of complex texts and real-world materials that extend beyond standard curriculum offerings. Such materials not only challenge students linguistically but also foster critical engagement with content, which is essential for their academic development.

However, designing tasks that meet these cognitive and linguistic demands presents a significant challenge for teachers. Crafting high-quality activities—such as inference-based comprehension questions, multi-layered listening tasks, or source-based reading exercises—requires both pedagogical expertise and considerable time investment. This challenge is further compounded by the need to continuously update materials to reflect current events, varied registers, and interdisciplinary topics, which are particularly relevant for gifted learners who thrive on novelty and real-world connections.

In recent years, the emergence of generative artificial intelligence (AI) tools has opened new possibilities for educational task design. Applications such as ChatGPT, Otter.ai, and other natural language processing (NLP) platforms enable educators to automate or accelerate key aspects of content creation—ranging from summarization and question generation to transcription and rephrasing. These tools offer not only efficiency but also the potential to tailor materials to diverse learner profiles, including the gifted. Yet, while promising, the integration of AI into pedagogical practice remains underexplored, particularly in relation to its application in designing high-level tasks for gifted students.

This study is therefore situated at the intersection of gifted education, language pedagogy, and AI-assisted instructional design. It seeks to examine how AI can be used meaningfully to support teachers in creating engaging, high-cognitive-load reading and listening tasks, with a specific focus on gap-filling exercises from authentic news (listening) and inference-based comprehension formats such as True/False/Not Given tasks (reading).

2. Problem Statement and Research Novelty

While authentic materials—such as news articles, podcasts, and documentaries—are widely accessible and recognized as valuable resources for enriching the learning experience of gifted students, the process of transforming these materials into effective pedagogical tasks remains a significant challenge for teachers. Designing cognitively demanding activities that align with the advanced proficiency and analytical capabilities of gifted learners requires not only a deep understanding of task complexity, but also substantial time, effort, and instructional design expertise. In particular, tasks such as gap-filling based on authentic news or inference-heavy reading questions (e.g., True/False/Not Given, Locate Information) demand careful selection, adaptation, and validation to ensure both authenticity and educational value. Despite the rapid advancement of generative AI technologies—many of which offer functionalities relevant to educational content creation—there remains a paucity of empirical research investigating how

these tools can be meaningfully integrated into task design for gifted language learners. This gap underscores the need for a focused study that examines the practical benefits, challenges, and pedagogical implications of AI-assisted instructional design in the context of advanced language education.

3. Research Aim

This study aims to investigate how artificial intelligence (AI) applications can effectively support English language teachers in designing high-level reading and listening tasks specifically tailored for gifted students. The focus is placed on two task types commonly used to assess higher-order comprehension skills: gap-filling exercises derived from authentic news-based listening texts, and reading comprehension questions in the True/False/Not Given (T/F/NG) a. The study seeks not only to evaluate the usefulness of AI in generating these tasks, but also to establish a replicable workflow for each task type, identifying the sequential steps and corresponding AI tools used in the design process. The ultimate goal is to provide teachers with a practical AI-supported task design protocol that can serve as a foundation for prompt libraries and be adapted to develop additional item types in high-level assessments, particularly those used in gifted education or national competitions.

4. Research Questions

This study is guided by two central research questions that explore the practical integration of AI tools in the design of cognitively demanding language tasks for gifted learners

Question 1: What are the key steps and AI tools involved in teachers' design of gap-filling listening tasks using authentic news materials?

Question 2: What are the key steps and AI tools involved in teachers' design of True/False/Not Given reading tasks?

Chapter 2. Literature Review

This chapter reviews relevant theoretical and empirical literature to situate the study within current academic discourse. The focus is placed on how AI can be meaningfully adapted to support task design using authentic materials for gifted learners, addressing a notable gap in accessible, high-cognitive-load exercises.

1. Theoretical Foundations

1.1 Task-Based Language Teaching (TBLT)

Task-Based Language Teaching (TBLT) is a pedagogical approach that places meaningful, real-world communication at the heart of language learning. Rather than focusing primarily on isolated grammar structures or vocabulary drills, TBLT encourages learners to use language as a tool for achieving specific outcomes—such as solving a problem, completing a project, or understanding a real-life situation (Ellis, 2003). This emphasis on purpose-driven communication, combined with exposure to authentic input, makes TBLT particularly valuable for gifted learners, who tend to thrive on intellectually engaging and cognitively challenging tasks. One of the core strengths of TBLT lies in its flexibility: it allows teachers to incorporate complex, content-rich

materials—such as news articles, interviews, or podcasts—into the classroom in a way that mirrors how language is actually used in the world. These materials provide fertile ground for designing advanced reading and listening tasks that go beyond surface-level comprehension, fostering critical thinking, inference, and synthesis—skills essential for the academic growth of high-ability students (Nunan, 2004).

1.2. Cognitive Load and Gifted Education

Gifted students benefit most from tasks that challenge their cognitive abilities and promote deeper thinking. Cognitive Load Theory (Sweller, 1988) suggests that learning is most effective when instructional materials balance mental effort with opportunities to build and apply knowledge. Because gifted learners typically process information quickly and efficiently, overly simple or repetitive tasks can lead to boredom and disengagement. Instead, they need tasks that introduce germane cognitive load, such as those involving problem-solving, abstract reasoning, and critical analysis (Van Merriënboer & Sweller, 2005). When paired with differentiated instruction strategies—like tiered assignments or flexible content—these tasks can be tailored to maintain challenge and foster meaningful learning, especially in language education where advanced comprehension and inference are key.

1.3. Bloom's Taxonomy and Higher-Order Comprehension

For C1–C2 level learners, language tasks should move beyond basic recall to foster deeper, more analytical engagement with texts. The revised Bloom's Taxonomy offers a practical framework for designing activities that target higher-order thinking skills such as analysis, evaluation, and creation (Anderson & Krathwohl, 2001). These skills are especially valuable in reading and listening tasks that involve inference, synthesis, or critical judgment. Gifted students, in particular, benefit from such cognitively rich tasks, as they encourage engagement with complexity, ambiguity, and multiple perspectives—features common in authentic materials like editorials or news reports. By applying Bloom's taxonomy to task design, educators can promote deeper comprehension and reflective language use (Forehand, 2010).

2. AI in Language Education

2.1. AI Applications in Designing Language Tasks for Gifted Learners

Recent developments in artificial intelligence have created new opportunities for designing advanced language tasks, especially for gifted learners at the C1–C2 level. These students benefit from activities that promote analysis, synthesis, and critical thinking. Tools like ChatGPT, NotebookLM, Whisper, and Quillbot support teachers by generating questions, simplifying complex texts, and adapting authentic materials efficiently. Used thoughtfully, these tools not only save time but also help create tasks that remain authentic and cognitively challenging. As Zawacki-Richter et al. (2019) note, AI can enhance both instructional quality and teacher productivity when integrated with clear pedagogical goals.

2.2. Generative AI and Natural Language Processing (NLP) in Task Design

Generative AI and NLP tools are becoming valuable assets in language education, especially for designing tasks that challenge advanced learners. Platforms like ChatGPT, Whisper, Quillbot, and NotebookLM assist teachers in generating questions, simplifying texts, transcribing

audio, and adapting content to learner levels. For gifted students at the C1–C2 level, these tools offer efficient ways to create tasks that promote inference, synthesis, and critical thinking. When used thoughtfully, AI supports authentic, engaging instruction without replacing the teacher's role. As Luckin et al. (2016) emphasize, its strength lies in enhancing—not substituting—pedagogical design.

2.3. AI for Adapting Authentic Materials into Pedagogical Tasks

A key challenge in advanced language education is turning authentic materials—like news articles, interviews, or documentaries—into structured, level-appropriate tasks. While such content offers rich input for gifted learners at the C1–C2 level, it often demands significant time and effort to adapt. AI tools now make this process more efficient. For example, NotebookLM can generate inference-based questions from uploaded texts, and ChatGPT can turn transcribed audio into targeted gap-filling exercises. These tools help teachers create differentiated, cognitively demanding tasks more easily. As Popenici and Kerr (2017) note, AI can act as a valuable co-designer in transforming real-world content into effective learning experiences.

2.4. Pedagogical and Practical Challenges in AI Integration

While AI offers exciting possibilities for language education, its use also raises important concerns. The quality of AI-generated content can vary, with issues like factual errors, unclear phrasing, or mismatched difficulty—particularly problematic for gifted learners who need precise, cognitively challenging tasks. Overreliance on AI may also reduce teachers' active role in task design. In addition, AI systems can reflect bias from their training data, affecting the fairness of materials (Bender et al., 2021). Many educators also lack training in AI literacy and prompt design, making it harder to use these tools effectively. These challenges underscore the need for thoughtful human oversight to ensure tasks remain accurate, appropriate, and pedagogically sound.

2.5. Gaps in Existing Research

Although interest in AI in education is growing, research on its use for designing tasks specifically for gifted language learners remains limited. Most existing studies focus on general applications of AI in language learning or assessment, with little attention to the unique cognitive and linguistic needs of high-ability students. While tools like ChatGPT and NotebookLM are increasingly used to generate questions or adapt texts, there is still a lack of clear, documented workflows that help teachers meaningfully integrate AI into instructional design—especially when working with authentic materials. As Holmes et al. (2019) note, AI in education is still developing, and there is a pressing need for pedagogically grounded frameworks that align AI use with differentiated instruction and the learning profiles of advanced learners. Filling this gap is essential to ensure that AI enhances, rather than dilutes, the quality of instruction for gifted students.

3. Needs Analysis

3.1. Definition and Importance of Needs Analysis

Needs analysis is vital in developing language curricula, ensuring alignment with learners' goals and proficiency levels. It identifies students' necessities, lacks, and wants, facilitating the creation of tailored syllabi that enhance advanced listening and speaking skills for English-majored students preparing for the NCGS. This approach enables educators to craft curricula that challenge

students' critical thinking and communication skills, essential for success in high-stakes environments like the NCGS.

3.2. Target Needs Analysis

In language education, target needs refer to specific skills required for particular communicative contexts, distinct from general language requirements (Hutchinson & Waters, 1987). For gifted students, target needs align with advanced language demands, such as excelling in competitions like the NCGS. These needs include necessities, essential competencies like advanced listening and speaking for academic debates (Nunan, 1988); lacks, gaps in proficiency hindering goals (Allwright & Allwright, 1977); and wants, personal aspirations like mastering public speaking (Hutchinson & Waters, 1987). Addressing these ensures targeted interventions, enhances motivation, and creates a curriculum aligned with students' ambitions, fostering active participation and skill development.

Chapter 3. Methodology

1. Research Participants

This study involved a purposive sample of six English language teachers from a gifted school, all of whom had direct experience teaching high-achieving students at the upper-secondary level. These teachers were selected based on their active involvement in curriculum development, enrichment programs, and task design for gifted learners. Each participant had a minimum of six years of teaching experience and demonstrated familiarity with advanced language instruction. In addition, the study included fifteen gifted students from Grades 11 and 12 at the same institution. All student participants were identified as high performers based on their academic records and formal inclusion in national or school-level gifted programs. Their English proficiency was confirmed to be at the C1 level of the CEFR through institutional placement tests and ongoing classroom assessments, ensuring their suitability for the cognitively demanding tasks involved in this research.

2. Research Questions and Design

This research adopts a qualitative case study approach, supported by a needs analysis framework. The case study enables in-depth exploration of how teachers engage with AI tools to design high-level language tasks, while the needs analysis identifies the specific instructional gaps and challenges teachers face in this process.

The study is guided by the following research questions:

Question 1: What are the key steps and AI tools involved in teachers' design of gap-filling listening tasks using authentic news materials?

Question 2: What are the key steps and AI tools involved in teachers' design of True/False/Not Given reading tasks?

3. Data Instruments

To support the design and evaluation of language tasks for gifted learners, this study employed a combination of AI applications and qualitative tools, selected for their ability to help teachers transform authentic materials into structured, cognitively demanding exercises. ChatGPT

was used to generate comprehension questions, paraphrase spoken or written texts, revise sentence structures, correct grammar, and synthesize content from multiple sources. Otter.ai and Whisper enabled the transcription of real-world audio content, such as news reports, into editable text for listening task development. Quillbot and TextCortex assisted in simplifying and rephrasing complex input, allowing teachers to adapt authentic texts without compromising their linguistic richness. NotebookLM was used to extract content from online sources, generate video scripts, identify challenging vocabulary, and create gap-filling tasks aligned with C1–C2 level descriptors. Finally, Play.ht was employed to convert written text into natural-sounding audio, supporting the development of realistic listening input. Together, these tools allowed for the efficient creation of high-level reading and listening tasks that align with the linguistic and cognitive needs of gifted students.

In addition to AI applications, several qualitative instruments were used to gather in-depth data on participants' experiences and perceptions. Pre- and post-task surveys and semi-structured interviews with teachers were conducted to explore their instructional needs, challenges in task design, and reflections on the effectiveness and usability of AI tools. These instruments captured both their initial expectations and their evolving views throughout the study. To gain insight from the learner perspective, student feedback forms were used to evaluate the clarity, difficulty, and engagement level of both AI-assisted and manually designed tasks. Finally, a collection of task samples—including both teacher-generated and AI-supported versions—was analyzed to compare quality, alignment with CEFR descriptors, and cognitive demand. This comparative analysis provided a clearer picture of the instructional value and practical potential of AI-assisted task design for gifted language learners.

4. Listening Task Design Model

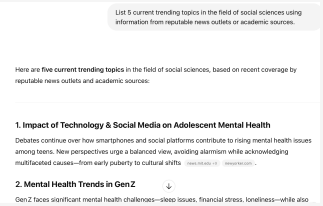
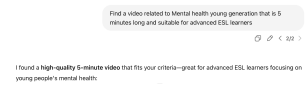
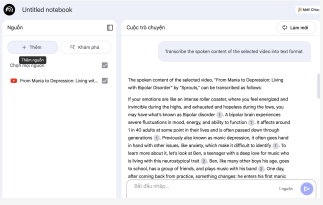
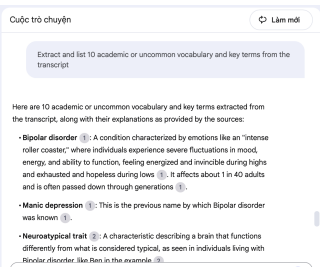
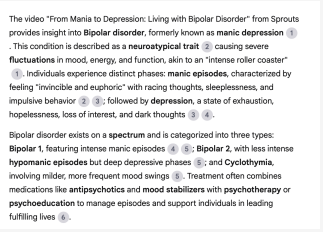
4.1. Design Gap-filling Listening Exercise from Authentic Materials

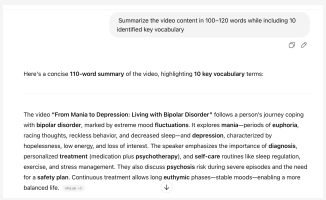
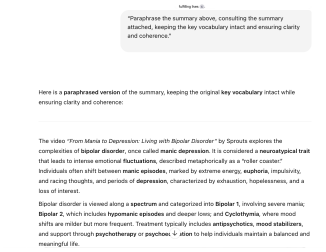
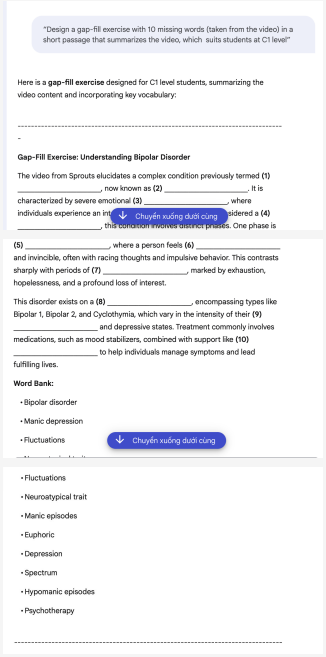
4.1.1. General Instruction

Step	Task	Outcome	Tool	Note
1	Compile trending topics in the field of social sciences from credible sources	A list of timely, academic-related topics	ChatGPT	Include a prompt asking for reputable sources (e.g., BBC, The Guardian)
2	Search for a suitable video based on topic and duration (4–6 minutes)	A video matching content and time requirements	ChatGPT or Gemini	Video quantity may vary according to the teacher's intended use
3	Transcribe the video content	Accurate transcript	Otter.ai / Whisper / NotebookLM	Review the transcript for accuracy if generated by AI

4	Identify difficult vocabulary or key terms from the transcript	A list of key or challenging vocabulary	NotebookLM	Can be categorized by word type (nouns, verbs, collocations, etc.)
5	Summarize the video content while retaining key vocabulary	A concise summary containing key terms	ChatGPT or NotebookLM	Ensure the summary reflects the core message of the video
6	Paraphrase the summary while keeping the key words	A paraphrased version that is clear and level-appropriate	ChatGPT or NotebookLM	Adapt the language to suit C1–C2 learners
7	Create a gap-fill listening exercise based on the paraphrased summary	A gap-fill task with answer key	ChatGPT or NotebookLM	Include around 5 - 10 blanks using selected key terms

4.1.2. Prompt and Illustration

Step	Task	Prompt	Outcome	Illustration
1	Compile trending topics in the field of social sciences from credible sources (ChatGPT)	"List 5 current trending topics in the field of social sciences using information from reputable news outlets or academic sources."	A list of 5 relevant and timely social science topics for lesson planning.	
2	Search for a suitable video based on topic and duration (4–6 minutes) (ChatGPT)	"Find a video related to [selected topic] that is 4–6 minutes long and suitable for advanced ESL learners."	A selected video that aligns with topic and timing criteria, ready for transcription.	
3	Transcribe the video content (NotebookLM)	"Transcribe the spoken content of the selected video into text format."	Accurate transcript of the video in text format for further processing.	
4	Identify difficult vocabulary or key terms from the transcript (NotebookLM)	Extract and list 10 academic or uncommon vocabulary and key terms from the transcript.	A list of approximately 10 difficult or subject-specific terms	
5	Summarize the video content while retaining key vocabulary (ChatGPT + NotebookLM)	Summarize the video content in 100–120 words while including 10 identified key vocabulary	A condensed version of the video with key terms, suitable for academic listening.	

				
6	Paraphrase the summary while keeping the key words (ChatGPT + NotebookLM)	“Paraphrase the summary above, consulting the summary attached, keeping the key vocabulary intact and ensuring clarity and coherence.”	A paraphrased summary maintaining terminology consistency and simplified structure.	
7	Create a gap-fill listening exercise based on the paraphrased summary (NotebookLM)	“Design a gap-fill exercise with 10 missing words (taken from the video) in a short passage that summarizes the video, which suits students at C1 level”	A complete gap-fill exercise with answer key, aligned with C1–C2 listening objectives.	

4.1.3. Exercises

a. Tape-script and Vocabulary List

The spoken content of the selected video, "From Mania to Depression: Living with Bipolar Disorder" by "Sprouts," can be transcribed as follows:

“If your emotions are like an intense roller coaster, where you feel energized and invincible during the highs, and exhausted and hopeless during the lows, you may have what’s known as Bipolar disorder. A bipolar brain experiences severe fluctuations in mood, energy, and ability to function. It affects around 1 in 40 adults at some point in their lives and is often passed down through generations. Previously also known as manic depression, it often goes hand in hand with other issues, like anxiety, which make it difficult to identify. To learn more about it, let’s look at Ben, a teenager with a deep love for music who is living with this neuroatypical trait. Ben, like

many other boys his age, goes to school, has a group of friends, and plays music with his band. One day, after coming back from practice, something changes: he enters his first manic episode. During mania, Ben feels invincible and euphoric without a particular reason. Thousands of thoughts are racing through his mind, and just as many future plans—not all of them realistic. He works tirelessly for days and hardly gets any sleep. Full of excitement he calls his friends and talks for hours, switching topics so fast that even he can't keep up. After Ben impulsively buys lots of stuff with his mother's credit card, she confronts him, but instead of being remorseful, he gets furious. Then, after a week, everything suddenly changes. Ben enters depression. Now he feels exhausted and lays in bed all day. When he does get up he finds nothing worth living for. Even the simple thought of getting dressed seems meaningless. He also has no appetite, and often skips meals.

Since Ben has also lost all interest in music, he doesn't show up for practice and later he feels guilty about it—telling himself he's a worthless friend. This is when he feels the lowest and starts to have dark thoughts. The symptoms of bipolar disorder look different in each case, which is why the condition is so difficult to diagnose and like all mental disorders, can only be defined on a spectrum. Traditionally, three types are recognized: Bipolar 1, Bipolar 2, and Cyclothymia. Bipolar 1 is characterized by intense manic episodes that can last a week or more, and sometimes ends with a trip to the hospital due to risky decisions. These are followed by a depressive episode. Bipolar 2 is characterized by less intense hypomanic episodes that can last 3 or 4 days, but deep depressive phases with dark ideas and self-harm. Cyclothymia is characterized by mood swings between episodes of hypomania and mild depression. While emotional changes of Bipolar 1 and 2 can spread over weeks, the ups and downs of cyclothymia often happen spontaneously over shorter periods—even within the same day.

The condition is often treated with antipsychotics, and mood stabilizers, that dampen the intensity and frequency of the episodes. The medications are often combined with psychotherapy or psychoeducation to help stabilize daily activities and reduce substance abuse and relapse. People like Ben can often feel helpless when they are first diagnosed, but eventually may realize that they can live fulfilling lives, have families, and have successful careers. Some feel as if their atypical brain makes them more creative, and choose to pursue a career in the arts. What do you know about bipolar? Do you have, or know someone with the disorder? If so, what do you know about supporting those coping with it? Tell us your experiences and thoughts in the comments below! We relentlessly researched education for the past decade, read plenty of peer-reviewed papers, visited all sorts of schools, created hundreds of videos on the topic, and even set up our very own kindergarten for project-based learning. And now, we also made a book!

The Unschoolers Educational Dictionary covers everything we know about learning, unveils the unintended consequences of compulsory education, and includes a complete guide to the alternative school movement. Order your copy now. Just visit your favorite book shop or check the links in the description below. If you found this helpful, check out our other videos, join us on youtube or on patreon.com/sprouts. For more information, classroom exercises or videos without background music, visit sproutsschools.com. Published under creative commons."

Vocabulary List:

• **Bipolar disorder**: A condition where emotions are like an intense roller coaster, involving severe fluctuations in mood, energy, and ability to function. It was previously known as manic depression.

• **Manic depression**: The previous name for Bipolar disorder.

• **Fluctuations**: Refers to the severe changes experienced in mood, energy, and ability to function by a bipolar brain.

• **Neuroatypical trait**: Describes a characteristic or feature of a brain that functions differently from what is considered typical, as exemplified by Ben who is living with bipolar disorder.

• **Manic episode**: A period during which a person feels energized, invincible, and euphoric without a particular reason. During such an episode, thousands of thoughts and future plans (not always realistic) race through the mind, and the individual may work tirelessly with little sleep. Impulsive actions, like excessive talking or spending, can also occur.

• **Euphoric**: A state of intense happiness and excitement, characteristic of a manic episode.

• **Impulsively buys**: Refers to buying things without careful thought or planning, a behavior observed in manic episodes.

• **Depression**: A phase that follows a manic episode, where a person feels exhausted, finds nothing worth living for, lacks appetite, and loses interest in activities. During this phase, individuals may feel worthless and have dark thoughts.

• **Spectrum**: Indicates that the symptoms of bipolar disorder can look different in each case and that, like all mental disorders, it can only be defined across a range of severities and presentations.

• **Bipolar 1**: A type of bipolar disorder characterized by intense manic episodes that can last a week or more, sometimes leading to hospitalization due to risky decisions, followed by a depressive episode.

• **Bipolar 2**: A type of bipolar disorder characterized by less intense hypomanic episodes that can last 3 or 4 days, but includes deep depressive phases with dark ideas and self-harm.

• **Cyclothymia**: A type of bipolar disorder characterized by mood swings between episodes of hypomania and mild depression, where these ups and downs often happen spontaneously over shorter periods, even within the same day.

• **Hypomanic episodes**: Less intense manic episodes, lasting 3 or 4 days, characteristic of Bipolar 2 and Cyclothymia.

• **Antipsychotics**: A type of medication used to treat bipolar disorder, often alongside mood stabilizers, to dampen the intensity and frequency of episodes.

• **Mood stabilizers**: Medications used to treat bipolar disorder that help dampen the intensity and frequency of episodes.

• **Psychotherapy**: A form of treatment often combined with medication to help stabilize daily activities and reduce substance abuse and relapse in individuals with bipolar disorder.

• **Psychoeducation**: A treatment approach often combined with medication that aims to educate individuals about their condition, helping to stabilize daily activities and reduce substance abuse and relapse.
• **Substance abuse and relapse**: Issues that medications combined with psychotherapy or psychoeducation aim to reduce in individuals with bipolar disorder.
• **Unschoolers Educational Dictionary**: The title of a book created by "Sprouts" that covers learning, the unintended consequences of compulsory education, and a guide to the alternative school movement.

b. Exercise

Gap-Fill Exercise: Understanding Bipolar Disorder

The video from Sprouts elucidates a complex condition previously termed **(1)** _____, now known as **(2)** _____. It is characterized by severe emotional **(3)** _____, where individuals experience an intense "roller coaster" of moods. Considered a **(4)** _____, this condition involves distinct phases. One phase is **(5)** _____, where a person feels **(6)** _____ and invincible, often with racing thoughts and impulsive behavior. This contrasts sharply with periods of **(7)** _____, marked by exhaustion, hopelessness, and a profound loss of interest.

This disorder exists on a **(8)** _____, encompassing types like Bipolar 1, Bipolar 2, and Cyclothymia, which vary in the intensity of their **(9)** _____ and depressive states. Treatment commonly involves medications, such as mood stabilizers, combined with support like **(10)** _____ to help individuals manage symptoms and lead fulfilling lives.

Answer Key: Understanding Bipolar Disorder

1. Manic depression
2. Bipolar disorder
3. Fluctuations
4. Neuroatypical trait
5. Manic episodes
6. Euphoric
7. Depression
8. Spectrum
9. Hypomanic episodes
10. Psychotherapy

4.2. Feedback from Teachers and Students

Teacher feedback on the use of AI tools to design gap-filling listening exercises for gifted students preparing for academic competitions reveals both notable strengths and important limitations. On the one hand, educators highlight the efficiency gains afforded by AI, particularly in automating labor-intensive tasks such as audio transcription, script generation, and the creation of gap-fill items. This enables teachers to devote more time to personalized instruction

and formative assessment. Moreover, when prompted effectively, AI can generate materials of high linguistic quality, closely aligned with C1–C2 proficiency standards. These materials often incorporate complex syntax, advanced vocabulary, and authentic language use—features that are critical for engaging and challenging high-achieving learners. On the other hand, a significant limitation lies in AI's lack of pedagogical discernment. Without the guidance of an experienced educator, AI-generated tasks may fail to address key elements of listening comprehension, including inferencing, prosodic features, and appropriate cognitive load in item design. Therefore, while AI serves as a powerful aid in instructional design, its outputs must be carefully curated and adapted to meet the specific educational aims and cognitive demands of advanced learners.

The use of AI-designed gap-filling listening exercises in preparation for the national contest for gifted students in English has demonstrated several pedagogical benefits. These tools enable more frequent engagement with advanced-level materials, thereby enhancing listening stamina and increasing exposure to complex linguistic input. Additionally, the wide range of topics covered—spanning domains such as science, education, and culture—contributes to the development of content familiarity and test readiness. Nonetheless, certain limitations were observed by students. Some exercises focused predominantly on surface-level word recognition rather than deeper comprehension skills, and the AI-generated audio occasionally lacked natural intonation and authentic pacing. While such tools provide valuable supplementary practice, their effectiveness is significantly improved when tasks are reviewed and refined by experienced educators to ensure alignment with the cognitive and linguistic demands of high-stakes competition contexts.

5. Reading Task Design Model

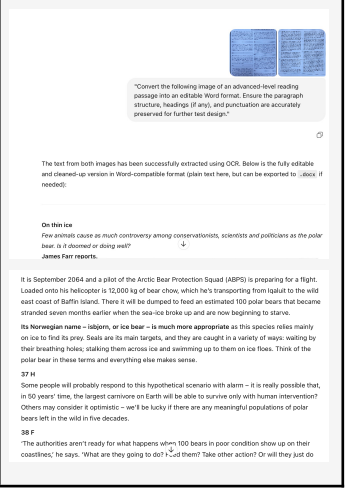
5.1. Design True/False/Not Given Reading Task

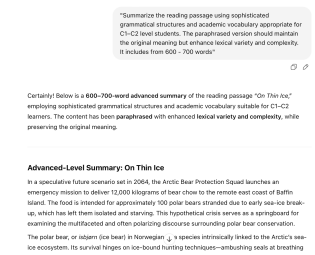
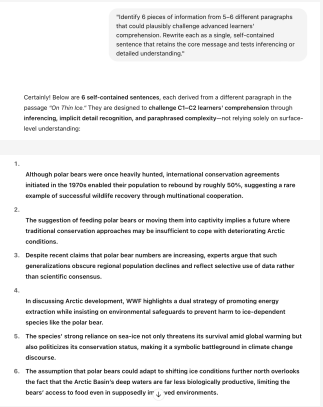
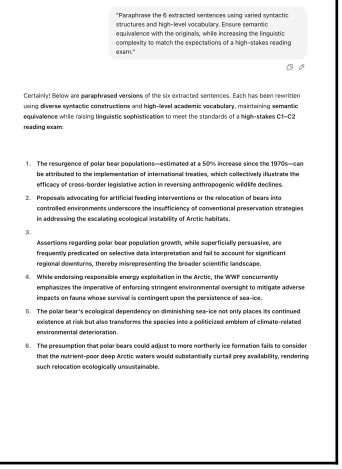
5.1.1. General Instruction

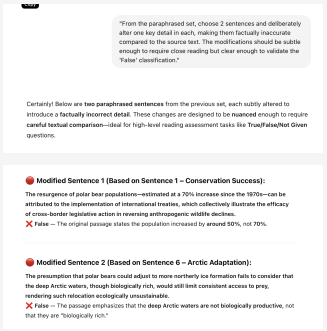
Step	Task	Outcome	Tool	Note
1	Select a reading passage from CAE/CPE books or online articles, take a photo, and convert it to Word format	Reading passage in Word format	Teacher + ChatGPT/Gemini	Input image should be clear
2	(Optional) Paraphrase the passage	Rephrased version of the original text	ChatGPT/ Gemini	Can be skipped if not necessary
3	Select 6 key pieces of information from 5–6 paragraphs that are likely to confuse students, and rewrite	6 simple sentences with core information	Teachers + ChatGPT/Gemini	Content taken from different paragraphs

	them as 6 individual sentences			
4	Paraphrase the 6 sentences	6 reworded sentences with equivalent meaning	ChatGPT/ Gemini / Quillbot / TextCortex	Use paraphrasing to avoid exact matches with the original text while maintaining a C1–C2 level of language complexity
5	Choose 2 sentences and edit one detail to make them incorrect compared to the passage → 2 “False” answers	2 sentences containing incorrect information	Teachers + ChatGPT/Gemini	Only change one small detail to make the sentence clearly False
6	Choose 2 sentences and add information **not found** in the passage → 2 “Not Given” answers	2 sentences containing information not present in the passage	Teachers + ChatGPT/Gemini	Add info that is not mentioned or implied in the original passage

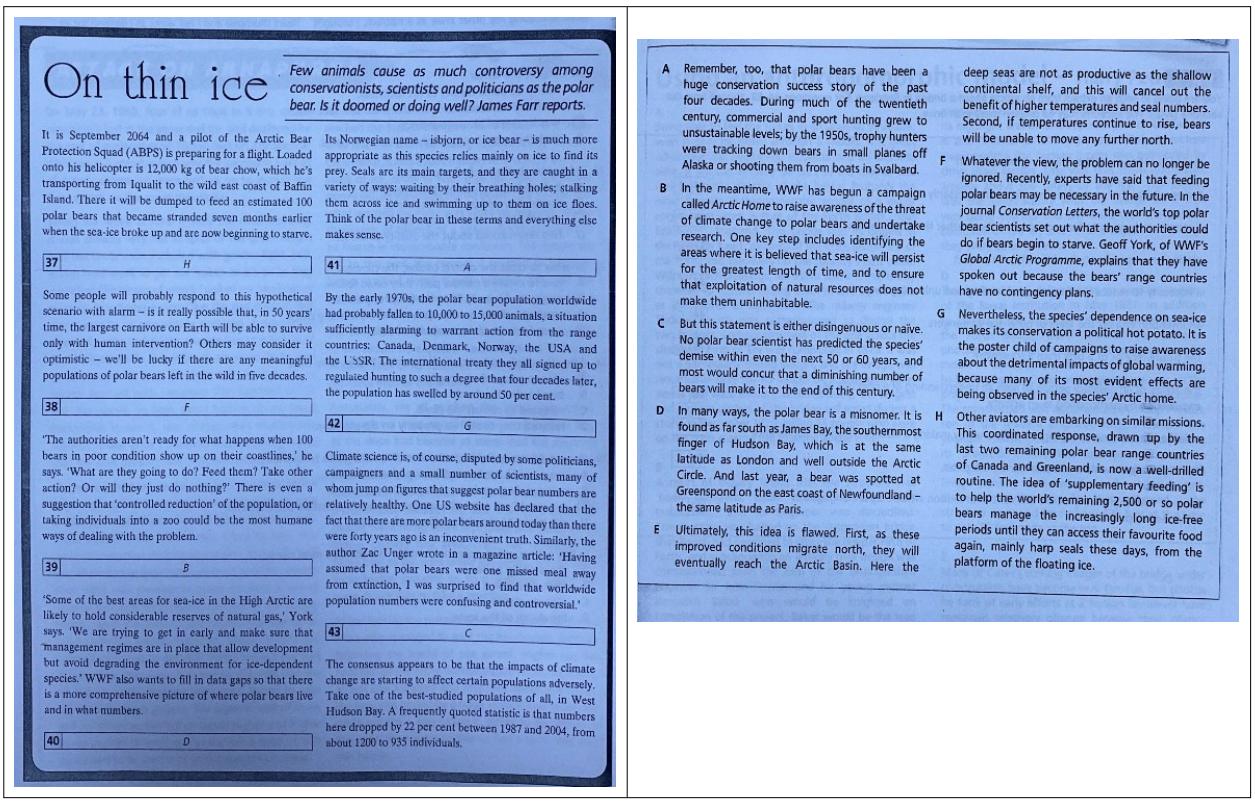
5.1.2. Prompt and Illustration

Step	Task	Prompt	Outcome	Illustration
1	Select a reading passage from CAE/CPE books or online articles, take a photo, and convert it to Word format	"Convert the following image of an advanced-level reading passage into an editable Word format. Ensure the paragraph structure, headings (if any), and punctuation are accurately preserved for further test design."	Reading passage in Word format	

2	(Optional) Paraphrase / Summarize the passage	"(Optional) Paraphrase/Summarize the reading passage using sophisticated grammatical structures and academic vocabulary appropriate for C1–C2 level students. The paraphrased version should maintain the original meaning but enhance lexical variety and complexity."	Rephrased/ Summarized version of the original text	
3	Select 6 key pieces of information from 5–6 paragraphs that are likely to confuse students, and rewrite them as 6 individual sentences	"Identify 6 pieces of information from 5–6 different paragraphs that could plausibly challenge advanced learners' comprehension. Rewrite each as a single, self-contained sentence that retains the core message and tests inferencing or detailed understanding."	6 simple sentences with core information	
4	Paraphrase the 6 sentences	"Paraphrase the 6 extracted sentences using varied syntactic structures and high-level vocabulary. Ensure semantic equivalence with the originals, while increasing the linguistic complexity to match the expectations of a high-stakes reading exam."	6 reworded sentences with equivalent meaning	

5	Choose 2 sentences and edit one detail to make them incorrect compared to the passage → 2 “False” answers	"From the paraphrased set, choose 2 sentences and deliberately alter one key detail in each, making them factually inaccurate compared to the source text. The modifications should be subtle enough to require close reading but clear enough to validate the 'False' classification."	2 sentences containing incorrect information	
6	Choose 2 sentences and add information not found in the passage → 2 “Not Given” answers	"From the same paraphrased set, select 2 sentences and insert information that is absent or not inferable from the original text. The added content should be contextually plausible, ensuring the sentence remains coherent, but must meet the standard for a 'Not Given' response."	2 sentences containing information not present in the passage	

5.1.3. Exercises
a. Original Text



On thin ice

Few animals cause as much controversy among conservationists, scientists and politicians as the polar bear. Is it doomed or doing well? James Farr reports.

It is September 2064 and a pilot of the Arctic Bear Protection Squad (ABPS) is preparing for a flight. Loaded onto his helicopter is 12,000 kg of bear chow, which he's transporting from Iqualit to the wild east coast of Baffin Island. There it will be dumped to feed an estimated 100 polar bears that became stranded seven months earlier when the sea-ice broke up and are now beginning to starve.

37 H

Some people will probably respond to this hypothetical scenario with alarm – is it really possible that, in 50 years' time, the largest carnivore on Earth will be able to survive only with human intervention? Others may consider it optimistic – we'll be lucky if there are any meaningful populations of polar bears left in the wild in five decades.

38 F

'The authorities aren't ready for what happens when 100 bears in poor condition show up on their coastlines,' he says. 'What are they going to do? Feed them? Take other action? Or will they just do nothing?' There is even a suggestion that 'controlled reduction' of the population, or taking individuals into a zoo could be the most humane ways of dealing with the problem.

39 B

'Some of the best areas for sea-ice in the High Arctic are likely to hold considerable reserves of natural gas,' York says. 'We are trying to get in early and make sure that management regimes are in place that allow development but avoid degrading the environment for ice-dependent species.' WWF also wants to fill in data gaps so that there is a more comprehensive picture of where polar bears live and in what numbers.

40 D

Its Norwegian name – isbjorn, or ice bear – is much more appropriate as this species relies mainly on ice to find its prey. Seals are its main targets, and they are caught in a variety of ways: waiting by their breathing holes; stalking them across ice and swimming up to them on ice floes. Think of the polar bear in these terms and everything else makes sense.

41 A

By the early 1970s, the polar bear population worldwide had probably fallen to 10,000 to 15,000 animals, a situation sufficiently alarming to warrant action from the range countries: Canada, Denmark, Norway, the USA and the USSR. The international treaty they all signed up to regulated hunting to such a degree that four decades later, the population has swelled by around 50 per cent.

42 G

Climate science is, of course, disputed by some politicians, campaigners and a small number of scientists, many of whom jump on figures that suggest polar bear numbers are relatively healthy. One US website has declared that the fact that there are more polar bears around today than there were forty years ago is an inconvenient truth. Similarly, the author Zac Unger wrote in a magazine article: 'Having assumed that polar bears were one missed meal away from extinction, I was surprised to find that worldwide population numbers were confusing and controversial.'

43 C

The consensus appears to be that the impacts of climate change are starting to affect certain populations adversely. Take one of the best-studied populations of all, in West Hudson Bay. A frequently quoted statistic is that numbers here dropped by 22 per cent between 1987 and 2004, from about 1200 to 935 individuals.

A Remember, too, that polar bears have been a huge conservation success story of the past four decades. During much of the twentieth century, commercial and sport hunting grew to unsustainable levels; by the 1950s, trophy hunters were tracking down bears in small planes off Alaska or shooting them from boats in Svalbard.

B In the meantime, WWF has begun a campaign called *Arctic Home* to raise awareness of the threat of climate change to polar bears and undertake research. One key step includes identifying the areas where it is believed that sea-ice will persist for the greatest length of time, and to ensure that exploitation of natural resources does not make them uninhabitable.

C But this statement is either disingenuous or naïve. No polar bear scientist has predicted the species' demise within even the next 50 or 60 years, and most would concur that a diminishing number of bears will make it to the end of this century.

D In many ways, the polar bear is a misnomer. It is found as far south as James Bay, the southernmost finger of Hudson Bay, which is at the same latitude as London and well outside the Arctic Circle. And last year, a bear was spotted at Greenspond on the east coast of Newfoundland – the same latitude as Paris.

E Ultimately, this idea is flawed. First, as these improved conditions migrate north, they will eventually reach the Arctic Basin. Here the deep seas are not as productive as the shallow continental shelf, and this will cancel out the benefit of higher temperatures and seal numbers. Second, if temperatures continue to rise, bears will be unable to move any further north.

F Whatever the view, the problem can no longer be ignored. Recently, experts have said that feeding polar bears may be necessary in the future. In the journal *Conservation Letters*, the world's top polar bear scientists set out what the authorities could do if bears begin to starve. Geoff York, of WWF's *Global Arctic Programme*, explains that they have spoken out because the bears' range countries have no contingency plans.

G Nevertheless, the species' dependence on sea-ice makes its conservation a political hot potato. It is the poster child of campaigns to raise awareness about the detrimental impacts of global warming, because many of its most evident effects are being observed in the species' Arctic home.

H Other aviators are embarking on similar missions. This coordinated response, drawn up by the last two remaining polar bear range countries of Canada and Greenland, is now a well-drilled routine. The idea of 'supplementary feeding' is to help the world's remaining 2,500 or so polar bears manage the increasingly long ice-free periods until they can access their favourite food again, mainly harp seals these days, from the platform of the floating ice.

b. Adapted Text

"On Thin Ice"

"In a speculative future scenario set in 2064, the Arctic Bear Protection Squad launches an emergency mission to deliver 12,000 kilograms of bear chow to the remote east coast of Baffin Island. The food is intended for approximately 100 polar bears stranded due to early sea-ice break-up, which has left them isolated and starving. This hypothetical crisis serves as a springboard for examining the multifaceted and often polarizing discourse surrounding polar bear conservation.

*The polar bear, or *isbjørn* (ice bear) in Norwegian, is a species intrinsically linked to the Arctic's sea-ice ecosystem. Its survival hinges on ice-bound hunting techniques—ambushing seals at breathing holes, stalking prey across vast expanses of frozen sea, or swimming to ice floes. Framing the species through this lens elucidates the existential threat it faces due to accelerating climate change and the resulting decline in sea-ice coverage.*

Concerns about the species' future elicit varying responses. Some experts adopt a pessimistic outlook, positing that polar bears may require human intervention to avoid extinction within five decades. Others perceive this projection as overly optimistic, suggesting that meaningful wild populations may not survive that long without drastic measures.

The urgency of intervention is underscored by recent expert discussions advocating potential feeding programs as contingency plans. However, this raises ethical dilemmas regarding whether allowing starvation, relocating bears to captivity, or undertaking controlled population reduction would be the most humane solution. Polar bear range countries, including Canada and Greenland, have implemented coordinated responses such as supplementary feeding routines,

wherein aerial missions deliver food to help the remaining 2,500 bears survive prolonged ice-free periods.

Despite these efforts, criticisms persist concerning global readiness to manage such scenarios. Environmentalists stress the need to identify and preserve areas where sea-ice will persist longest and to prevent unsustainable exploitation of natural resources. Simultaneously, conservationists like the WWF strive to close critical data gaps, seeking comprehensive knowledge on polar bear distribution and population dynamics.

Historically, polar bear numbers plummeted during the 20th century due to unregulated commercial and sport hunting. By the 1970s, the population had declined to between 10,000 and 15,000 individuals, prompting international agreements to curtail hunting. These efforts were largely successful; population numbers rebounded by approximately 50 percent over the following decades, positioning the polar bear as a symbol of conservation success.

Nevertheless, this recovery is uneven and contentious. Climate skeptics and some media outlets highlight increasing polar bear populations as evidence against global warming narratives. However, scientists argue that such claims are misleading. Although certain populations appear stable or growing, others—like those in West Hudson Bay—have shown marked decline, with numbers falling by 22 percent from 1987 to 2004.

The discourse is further complicated by political sensitivities. Polar bears have become emblematic of the broader climate crisis, often used to galvanize public and policy attention. This visibility, however, has made conservation a political "hot potato," entangled in debates over the legitimacy of climate science and appropriate governmental responses. Discrepancies in population data, shifting baselines, and inconsistent methodologies have exacerbated confusion, making consensus elusive.

Some experts remain skeptical of overly dire extinction forecasts, labeling them alarmist or unfounded. While no credible scientist claims the species will vanish within mere decades, there is widespread agreement that continued sea-ice loss and insufficient mitigation strategies will drive substantial declines. Even optimistic projections suggest the species will survive only in greatly reduced numbers unless global action is taken.

Additionally, assumptions about the bears' adaptive capacities are increasingly challenged. Warmer temperatures may improve ice conditions in some northern areas, yet this would only shift habitat distribution. Unfortunately, the Arctic Basin's deep-sea environments are far less biologically productive than the continental shelf, undermining any presumed advantage. As a result, migrating bears could struggle to find sufficient prey, exacerbating the survival crisis.

In conclusion, while polar bears have achieved a partial conservation comeback since the mid-20th century, their future remains precarious. The interplay of ecological dependency, political inertia, and scientific uncertainty renders their survival a litmus test for global environmental governance. Without proactive, internationally coordinated strategies, the species may soon exist only through human support, as living relics of a vanishing ice world."

c. Exercises and Answer Key

Exercises: True/False/Not Given

1. The resurgence of polar bear populations—estimated at a 70% increase since the 1970s—can be attributed to the implementation of international treaties, which collectively illustrate the efficacy of cross-border legislative action in reversing anthropogenic wildlife declines.
2. Proposals advocating for artificial feeding interventions or the relocation of bears into controlled environments underscore the insufficiency of conventional preservation strategies in addressing the escalating ecological instability of Arctic habitats.
3. The presumption that polar bears could adjust to more northerly ice formation fails to consider that the deep Arctic waters, though biologically rich, would still limit consistent access to prey, rendering such relocation ecologically unsustainable.
4. Assertions regarding polar bear population growth, while superficially persuasive, are frequently predicated on selective data interpretation and are often disseminated by oil industry lobbyists seeking to downplay the urgency of climate action.
5. While endorsing responsible energy exploitation in the Arctic, the WWF concurrently emphasizes the imperative of enforcing stringent environmental oversight and advocates for renewable energy investments to safeguard ice-dependent species.
6. The polar bear's ecological dependency on diminishing sea-ice not only places its continued existence at risk but also transforms the species into a politicized emblem of climate-related environmental deterioration.

Answer key:

1. **False** — The original passage states the population increased by **around 50%**, not **70%**.
2. **True**
3. **False** — The passage emphasizes that the deep Arctic waters are not biologically productive, not that they are "biologically rich."
4. **Not Given** — The original text discusses selective use of population data but makes no mention of oil industry involvement or motive.
5. **Not Given** — Although the text references resource development regulation, it does not mention WWF support for renewable energy.
6. **True**

5.2. Feedbacks from Teachers and Students

Teacher feedback on the use of AI tools in designing True/False/Not Given (TFNG) reading exercises for gifted students preparing for high-stakes English competitions highlights both the benefits and limitations of this approach. A frequently cited strength is the efficiency with which AI can generate practice materials. By automating tasks such as text analysis and question construction, AI enables teachers to devote more time to higher-order pedagogical activities, including diagnostic feedback and strategy training. When prompted effectively, AI is also capable of producing TFNG items that meet the linguistic and cognitive demands of C1–C2 proficiency levels, demonstrating appropriate syntactic complexity and lexical sophistication. These tasks support the development of critical reading sub-skills, particularly in recognizing implied meaning and distinguishing factual statements from opinion. Nonetheless, concerns remain regarding the AI's limited pedagogical discernment. In particular, educators have noted that AI-generated items

sometimes fail to clearly differentiate between “False” and “Not Given,” potentially resulting in ambiguity and learner confusion. As such, careful teacher oversight is essential to ensure the validity, clarity, and appropriateness of AI-generated materials, especially in the context of competitive academic preparation.

Students working with AI-generated True/False/Not Given reading tasks often express mixed feedback regarding their experience. Many appreciate the variety and volume of practice materials made available in a short amount of time, which allows them to strengthen their reading speed and critical thinking under timed conditions. The AI-generated texts also tend to expose learners to a wide range of academic topics and complex structures, which is particularly useful for students aiming for C1–C2 level performance. However, some students report difficulties distinguishing between “False” and “Not Given” items, as the phrasing of the statements may occasionally lack clarity or seem misleading. In some cases, the questions test surface-level detail rather than deeper comprehension or inference, reducing their overall value as training tools. As a result, students often recommend that such exercises be reviewed or supplemented by teachers to ensure alignment with exam standards and to support more accurate interpretation.

Chapter 4. Findings and Conclusions

This study explored the integration of AI tools into the design of two task types targeting receptive skills: gap-filling listening exercises using authentic news materials and True/False/Not Given (TFNG) reading tasks based on advanced-level texts. A comparison was made between AI-generated and manually created tasks, focusing on efficiency, output quality, and learner engagement. Findings also address the two central research questions related to the task design process and tool selection.

1. AI vs. Manual Task Design

Across both listening and reading tasks, AI tools demonstrated clear advantages in terms of efficiency. The automation of time-consuming steps - such as transcribing audio, generating statements, and formatting items - significantly reduced the preparation workload. This was particularly useful for teachers working under time constraints or handling high volumes of materials for advanced learners.

Output quality, however, was more variable. When teachers provided specific, context-rich prompts - clearly outlining the task type, linguistic focus, number of items, proficiency level, and source material - AI-generated tasks were largely coherent, linguistically accurate, and aligned with the cognitive demands expected at C1–C2 level. In listening tasks, detailed prompts enabled the AI to identify blanks that tested key lexical or grammatical points rather than surface-level vocabulary. In reading tasks, clear distinctions between “False” and “Not Given” were more likely to be generated when the input guided the AI toward inference-based reasoning or paraphrased information.

In terms of learner engagement, students expressed interest in the diversity of AI-generated materials, particularly the use of real-world content. Authentic news clips and advanced-level articles exposed learners to rich input and current topics, increasing motivation and relevance.

However, some reported difficulties with item clarity or mismatches between the questions and source material, especially when AI outputs were not reviewed or refined by teachers.

2. The Role of Prompt Specificity

This study set out to develop a set of instructional guidelines to support teachers in designing advanced-level reading and listening tasks for gifted learners. It focuses on two task types commonly used to evaluate higher-order comprehension: **gap-filling exercises** based on authentic news listening texts and **True/False/Not Given (T/F/NG)** reading questions. The overarching aim is to offer a practical, AI-assisted framework for task creation—one that can serve as a foundation for building prompt libraries and be adapted for the development of additional high-level item types, particularly in the context of gifted education and national academic competitions.

To meed these goals, the study focused on two central research questions:

Question 1*: What are the key steps and AI tools involved in teachers' design of gap-filling listening tasks using authentic news materials?*

The design process began with selecting an authentic news video of suitable length (typically 3–6 minutes) and complexity for advanced learners. Teachers then used AI transcription tools such as Whisper or Otter.ai to produce accurate scripts. These were input into ChatGPT, which was prompted to identify linguistically relevant gaps—such as collocations, discourse markers, or key vocabulary. The final step involved teacher review to ensure that the selected gaps aligned with listening objectives (e.g., comprehension, word recognition) and maintained appropriate difficulty and distribution across the text.

Question 2*: What are the key steps and AI tools involved in teachers' design of True/False/Not Given reading tasks?*

Teachers selected advanced-level texts from sources such as CAE/CPE materials or academic articles. Using ChatGPT, they summarized paragraphs and extracted key ideas, then prompted the AI to generate 6–8 statements reflecting different levels of comprehension. Teachers modified some statements slightly to produce "False" items and added unrelated details for "Not Given" items. A final verification step ensured each item was clearly linked to the source text, free from ambiguity, and balanced in difficulty.

In both task types, the clarity and specificity of the AI prompt played a central role in determining output quality. Vague prompts often resulted in generic or inaccurate items, while well-structured prompts—specifying task type, skill focus, text type, and learner level—led to more coherent, accurate, and pedagogically useful results. These findings reinforce the importance of teacher expertise in both **prompt engineering** and post-generation review.

3. Limitations of AI Tools

While AI tools provided significant support in automating aspects of listening task creation —such as transcription and gap selection—they also presented several limitations. Most notably, AI lacked pedagogical sensitivity in determining which words or phrases to omit for gap-filling. In some cases, the selected blanks were either too predictable or linguistically insignificant, failing to challenge learners or target meaningful comprehension skills. Additionally, AI tools did not

account for listening-specific cognitive factors, such as speech rate, prosodic cues, or contextual inferencing, which are crucial in assessing listening comprehension. The audio quality and speaker variation—key elements in authentic listening practice—were also not addressed or evaluated by AI. These limitations highlight the need for careful teacher intervention to review, refine, and align AI-generated tasks with the intended learning outcomes.

Although AI tools proved useful in generating initial drafts of T/F/NG statements, several limitations were observed in their application. A key issue was the lack of precision in distinguishing between "False" and "Not Given" items, which often resulted in ambiguous or misleading statements. AI-generated content sometimes relied too heavily on surface-level rewording rather than deeper inferencing or paraphrasing, leading to items that did not accurately reflect higher-order comprehension demands. Moreover, the AI lacked the ability to evaluate textual coherence and distractor validity, occasionally producing statements that were either too obvious or too vague to be pedagogically effective. These shortcomings underscore the importance of teacher oversight to ensure that each item is clear, answerable, and appropriately aligned with the source text and assessment objectives.

4. Recommendations for Teacher–AI Collaboration

To ensure task quality and learning impact, the study recommends a teacher-AI partnership model. Teachers should take an active role in prompt design by clearly defining task goals, language level, content type, and target skills. AI should be seen as a drafting tool, with human oversight essential for final validation.

Importantly, all AI-generated tasks should be evaluated using a standardized rubric for task quality, assessing criteria such as linguistic accuracy, cognitive demand, alignment with proficiency descriptors, clarity of instructions, and distractor validity. This ensures not only the reliability of the materials but also supports professional reflection in task design.

In conclusion, while AI tools can significantly streamline task creation and diversify input materials, their effective use requires strategic teacher guidance. When supported by detailed prompts and rigorous quality checks, AI becomes a valuable asset in the development of engaging and pedagogically sound tasks for advanced learners.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In E. Hovy, M. Walker, & A. Stent (Eds.), *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery.

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

Forehand, M. (2010). Bloom's taxonomy. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology* (pp. 41–47). University of Georgia. https://textbookequity.org/ Textbooks/Orey_Emergin_Perspectives_Learning.pdf

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AI-in-Education-Promises-and- Implications.pdf

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed-Publication.pdf

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), 1–13. https://doi.org/10.1186/s41039-017-0062-8

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0 https://doi.org/10.1145/3442188.3445922