

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

TABLE OF CONTENTS	Pages
CHAPTER 1: INTRODUCTION	
1. Rationale of the Study	2
2. Objectives of the Study	2
3. Scope and Limitations	3
4. Structure of the Report	3
CHAPTER 2: DEVELOPMENT	
1. Literature Review	
1.1. AI in Language Education	4
1.2. AI Tools for Reading and Listening Material Design	5
1.3. Previous Studies on AI in Assessment	5
1.4. Gaps and Implications	5
2. Theoretical Background	
2.1. CEFR Framework and Advanced Learners (C1–C2)	6
2.2. Test Design Principles for Gifted Students	6
2.3. Cognitive Demand and Task Types in Listening & Reading	6
3. Methodology	
3.1. Data Collection Process	7
3.2. AI Tools Employed (Sider OCR, ChatGPT, Oxford Text Checker, etc.)	7
3.3. Design Workflow and Validation	7
3.4. Involvement of Teachers and Students	8
4. AI Integration in Practice	
4.1. Pre-processing Source Materials	8
4.2. Practical Example 1: Recycling a CPE Reading Task into a Summary Completion Listening Task	14
4.3. Practical Example 2: Recycling an Online Article into a Reading Cloze Test	20
5. Case Studies and Sample Tests	
5.1. Sample Listening Task: Recycled from Reading	24
5.2. Sample Task and Answer Key	24
5.3. Feedback from Teachers and Learners	25
5.4. Refinement Based on Feedback	26
6. Discussion and Evaluation	
6.1. Strengths and Challenges of AI Use	26
6.2. Reliability and Validity Concerns	27
6.3. Learner Autonomy and Test Familiarity	27
6.4. Implications for Future Test Design	27
CHAPTER 3: CONCLUSION AND RECOMMENDATION	
1. Summary of Key Findings	28
2. Practical Recommendations for Teachers	28
3. Suggestions for Further Research	29
REFERENCES	30
APPENDICES	
Appendix A: Recycled Listening Materials	
Appendix B: Recycled Reading Materials	
Appendix C: Quality Control	

CHAPTER 1. INTRODUCTION

1. Rationale of the study

In Vietnam, particularly in specialized high schools, teachers preparing students for the National Excellent Student Exam in English frequently utilize high-level exam preparation materials such as *Cambridge English: Proficiency (CPE)*, *Objective Proficiency*, *Expert Proficiency*, and other C1–C2 resources published by major international publishers. These materials are widely accessible and familiar to advanced students, many of whom actively seek them out and attempt to complete practice tests independently, often relying on published answer keys and detailed explanations. While this self-study approach can support skill development, it also poses challenges in terms of assessing genuine performance, since students may simply replicate the models or memorize answers without engaging in deeper language processing.

This reality highlights the essential role of teachers in carefully selecting and transforming input from these authentic materials. By curating high-level vocabulary, advanced grammar structures, and thematically rich content, teachers can recycle and reframe texts into customized tasks that both consolidate core knowledge and more accurately assess students' actual language competence. The ability to adapt and reinvent such materials is especially important for promoting critical thinking, avoiding overfamiliarity, and ensuring that test content remains both challenging and pedagogically purposeful.

In recent years, a wide range of AI tools have been developed for educational purposes, each offering specialized functions such as text generation, audio transcription, vocabulary simplification, or item analysis. However, the abundance of tools—with overlapping or limited capabilities—often creates confusion for teachers, who must determine how to combine them effectively when designing specific task types (e.g., multiple-choice inference questions in reading, or note-completion tasks in listening). Furthermore, most advanced AI applications operate on subscription-based models, requiring separate payments for full access. For teachers who rely on multiple tools across different phases of test construction, the accumulated cost can become a significant barrier.

To address these challenges, the author carefully curated a set of AI tools that either offer generous free usage tiers or allow for flexible short-term access. By prioritizing tools that balance functionality with affordability, the study presents a realistic, scalable model for teachers working under budget constraints. The paper focuses on answering the following questions:

1. How can the original formats of reading and listening materials be processed into editable, teacher-friendly test items?
2. How can AI tools support teachers in designing reading and listening tasks aligned with advanced academic standards?
3. How can teachers ensure the accuracy, appropriateness, and academic integrity of AI-generated test materials?

2. Objectives of the study

The goals behind this work lies in the belief that AI, when used thoughtfully and critically, can empower educators—not replace them. It can help bridge the gap between professional testing practices and the day-to-day realities of classroom instruction. This project is therefore both a reflection of the author's ongoing learning and a contribution to a growing community of teachers

who strive to make advanced English instruction more effective, manageable, and responsive to learners' needs.

3. Scope and Limitations of the study

❖ Scope of the study

This study explores the application of AI tools in designing advanced reading and listening tests for English majors at a Gifted High School, with a primary focus on the teacher's role in test construction. The scope involves the development of NEC-aligned materials through AI-powered platforms, enabling the creation of level-appropriate, skill-specific tasks that reflect the cognitive demands of national competitions.

During the design phase, the teacher utilized AI to generate, adapt, and calibrate reading passages and listening texts, ensuring alignment with CEFR C1–C2 descriptors. Emphasis was placed on integrating vocabulary control, syntactic complexity, and question types (e.g., inference, attitude, and reasoning items) commonly found in high-stakes assessments.

In the implementation phase, 12 advanced 11th-grade students engaged with the materials through practice sessions and mock tests. Their role included solving the tasks, reflecting on item quality, and providing feedback on content validity, clarity, and difficulty.

❖ Limitations

While the initial application of AI tools in designing reading and listening materials has yielded promising results, several limitations of the study should be acknowledged. Firstly, the number of test sets developed by the teacher remains limited, which may not fully capture the range of possible applications or variations in difficulty and design. Secondly, the integration of AI tools, though effective in many respects, may not have been optimized; there may exist more efficient or reliable combinations of tools, or alternative AI platforms that offer greater convenience and functionality. Additionally, this study focused exclusively on free-access AI features. As a result, the author has not yet explored the potential advantages of premium tools, which may offer enhanced features for test creation, customization, and evaluation. These constraints suggest that further experimentation—with a broader range of AI resources and more extensive test development—is necessary to fully understand and optimize the role of AI in designing high-level language assessments.

4. Structure of the report

The paper consists of three main chapters: Introduction, Development, and Conclusion.

Chapter I: introduction outlines the overall structure of the paper, presents the rationale for choosing the topic, defines the aims of the study, and describes the methodology.

Chapter II: development reviews key literature on AI applications in language assessment and outlines the practical process of designing AI-assisted reading and listening tests for gifted students. It highlights how AI enhances task authenticity, supports differentiated instruction, and strengthens teacher-led material development. The chapter also examines current classroom practices at a local gifted high school and identifies opportunities for AI integration.

Chapter III: conclusion evaluates the feasibility and pedagogical impact of integrating AI tools into designing listening and reading materials, emphasizing on the irreplaceability of teacher expertise, even though AI tools can assist with formatting and drafting tasks.

CHAPTER II – DEVELOPMENT

This chapter presents both the theoretical grounding and practical implementation of AI-supported test design, focusing on advanced-level reading and listening assessments tailored for gifted students. The chapter is divided into two main parts: a review of relevant literature and a description of the developmental process that guided the construction and classroom application of AI-generated tasks.

1. Literature Review

1.1. AI in Language Education

Reading and listening are foundational receptive skills that play a critical role in academic achievement, particularly for high-performing students preparing for competitive language examinations. According to Grabe and Stoller (2011), advanced reading comprehension extends far beyond decoding; it involves synthesizing information, drawing inferences, and engaging critically with complex texts. Similarly, Field (2008) highlights that proficient listening at higher levels requires the ability to process connected speech, recognize implied meanings, and retain information across extended spoken discourse.

However, as Richards (2015) points out, a significant challenge in language education lies in the limited availability of cognitively demanding and personalized materials designed specifically for gifted learners. This lack of appropriately leveled input often restricts opportunities for in-depth skill development, particularly in settings where academic excellence is a top priority.

1.2. AI Tools for Reading and Listening Material Design

In national contexts such as Vietnam, receptive skills form the backbone of high-stakes assessments like the National English Competition (NEC) and international qualifications such as the Cambridge English: Proficiency (CPE). The format of the 2024–2025 National Excellent Student Exam in English clearly reflects this emphasis: the listening section contributes 5 points and the reading section 8 points, accounting for 13 out of the total 20 points, or 65% of the overall score. This weighting underscores the dominant role of receptive skills in evaluating students' language proficiency and reinforces the need for rigorous, well-calibrated materials in these areas.

In response to these demands, recent advancements in artificial intelligence (AI) have introduced new possibilities in language assessment. Emerging studies suggest that AI can assist teachers in streamlining content creation, adapting tasks to suit learner profiles, and simulating authentic language input. Within the context of high-level tests like the NEC—where tasks require advanced inferencing, discourse comprehension, and syntactic sophistication—AI tools have shown particular promise in supporting task development.

The literature identifies several key benefits of AI integration:

- ❖ **Task authenticity:** AI can generate or adapt texts and audio based on real-world contexts, thereby enhancing the relevance of assessment input.

- ❖ **Differentiated instruction:** Teachers can leverage AI to tailor tasks to students' CEFR levels and individual learning needs.
- ❖ **Teacher agency:** AI functions as a support system, enabling educators to focus more on task refinement and pedagogical alignment.

1.3. Previous Studies on AI in Assessment

Recent studies have shown how AI is reshaping language assessment, especially in terms of automation, customization, and content alignment. Early work concentrated on automated scoring in writing and speaking, with tools like *e-rater* and *SpeechRater* proving useful but still requiring human review to address discourse-level nuances (Burstein et al., 2018).

Another research direction involves adaptive testing, where item difficulty adjusts in real time based on student responses. These systems improve efficiency and personalization, especially for mixed-ability or high-performing learners (Van der Linden, 2019).

AI has also been used to support item generation, particularly for reading and grammar tasks. Some studies explored how large language models can create multiple-choice questions and distractors based on authentic texts, though challenges remain in ensuring task validity and cognitive alignment (Zhong & Meurers, 2022).

In addition, researchers have applied AI tools to analyze CEFR levels of test inputs, enhancing the appropriateness and consistency of materials (Ellis & Winke, 2020). A growing trend emphasizes human-AI collaboration in test design, where AI assists with drafting and teachers refine content—a model suitable for advanced learners needing tailored, high-level input (Godwin-Jones, 2023).

1.4. Gaps and Implications

To contextualize these insights, the study examined current reading and listening instruction at a Gifted High School. Teachers reported spending significant time manually modifying existing test materials to match NEC standards, often without sufficient tools for lexical or grammatical analysis. Listening resources were especially limited in terms of available transcripts and suitable question types. This gap highlighted the potential for AI integration to support test development in both skill areas.

Nevertheless, researchers consistently caution against over-reliance on AI. While AI can aid in drafting and structuring materials, crucial aspects such as content validation, cultural sensitivity, and cognitive calibration still require human expertise. AI should therefore be seen as a tool to enhance teacher productivity—not a substitute for the professional judgment and instructional insight that educators bring to the design of high-quality language assessments.

2. Theoretical Background

2.1. CEFR Framework and Advanced Learners (C1–C2)

The Common European Framework of Reference for Languages (CEFR) provides a widely accepted scale for describing language proficiency. At the advanced levels (C1 and C2), learners are expected to exhibit near-native control of the language in terms of comprehension, production, and interaction. Specifically, C1 users can understand a wide range of demanding texts and recognize implicit

meaning, while C2 users can summarize information from different spoken and written sources, reconstructing arguments in a coherent presentation (Council of Europe, 2020).

For test design purposes, the CEFR serves as a guiding framework for calibrating task complexity, linguistic input, and expected output. Reading and listening tasks at these levels often involve dense academic texts, multi-layered arguments, or nuanced opinions. Moreover, tasks must assess not just surface comprehension but also inferencing, evaluation, and synthesis. Therefore, tests targeting C1–C2 learners must reflect this high level of cognitive and linguistic sophistication to remain valid and effective.

2.2. Test Design Principles for Gifted Students

Gifted students—particularly those in national or specialized programs—often require assessment tools that go beyond conventional difficulty scaling. According to Kaplan (2018), test design for gifted learners must include higher-order thinking, complex language exposure, and problem-solving in unfamiliar contexts. In language testing, this means tasks should require more than simple information retrieval; they should involve interpreting ambiguous input, evaluating contrasting perspectives, or manipulating abstract language.

Additionally, assessments for gifted learners should maintain cognitive stretch without sacrificing clarity of purpose. This entails a careful balance between task novelty and accessibility, as overly convoluted items may distort test validity. When integrating AI into the test design process, it is critical to preserve this alignment by ensuring that AI-generated content is both level-appropriate and pedagogically sound.

2.3. Cognitive Demand and Task Types in Listening & Reading

Reading and listening at the C1–C2 levels require learners to process dense input, non-linear structures, and implicit information. Alderson (2000) emphasizes that advanced receptive skills involve integrating content across paragraphs or speaker turns, decoding rhetorical devices, and interpreting tone and stance. As such, traditional question formats like multiple-choice or true/false may not fully capture a learner's comprehension depth unless paired with cognitively demanding content.

In listening, especially, learners at these levels must interpret extended discourse such as academic lectures, expert interviews, or panel discussions, often under time pressure. Task types such as summary completion, matching viewpoints, and inference-based MCQs are well-suited to such contexts (Buck, 2001). Similarly, reading tasks should include gap-filling with options, sentence insertion, and identifying writer's attitude or organizational structure—all of which reflect real-world reading purposes and challenge higher-order comprehension skills.

AI tools, when applied responsibly, can assist in creating such tasks by analyzing linguistic complexity, generating distractors, and checking CEFR alignment. However, their use must be grounded in psycholinguistic theory and test design principles to ensure meaningful assessment.

3. Methodology

3.1. Data Collection Process

The methodology employed in this study adopts a design-based research (DBR) approach, which combines theoretical exploration with iterative testing in real educational settings. The primary data sources included:

- Authentic reading passages and listening transcripts drawn from C1–C2 level materials (e.g., CPE, IELTS, TED Talks...);
- Student performance on pilot test items;
- Teacher observations and feedback;
- AI-generated output (e.g., distractors, question stems, grammar checks).

These sources were selected to ensure both linguistic richness and curricular relevance to gifted high school learners in Vietnam preparing for national or NEC-level competitions.

3.2. AI Tools Employed

The following AI-powered tools were integrated systematically at different stages of test development:

- ❖ **SIDER OCR:** Used to extract editable text from scanned books and PDF materials. This allowed efficient digitization of source materials while preserving layout and formatting.
- ❖ **ChatGPT:** Used for rephrasing, generating distractors, suggesting task types (e.g., gap-fill, MCQs, paraphrase-based cloze), and adapting texts to maintain CEFR alignment. Prompt engineering was used to guide the AI to produce pedagogically valid outputs.
- ❖ **Oxford Text Checker:** Employed to determine the CEFR level of vocabulary and grammatical structures in both source texts and test items, helping ensure consistency with C1–C2 descriptors.
- ❖ **TTSMaker:** Used to generate listening audio from adapted texts, providing native-like pronunciation and clarity suitable for test deployment.
- ❖ **Cathoven Level Adaptor:** used to adjust and enhance the lexico-grammatical complexity of texts, aligning them with targeted CEFR levels—particularly useful for upgrading materials for advanced learners.

These tools were selected for their accessibility, language precision, and ability to integrate into classroom workflows.

3.3. Design Workflow and Validation

The test design process followed a four-stage model:

1. **Source Text Selection:** Authentic texts were chosen based on thematic relevance, cognitive demand, and CEFR level.
2. **AI-Assisted Task Generation:** Prompts were used to instruct ChatGPT to generate question types (summary completion, sentence insertion, etc.) aligned with the chosen text.
3. **Teacher Review and Editing:** Generated items were checked for appropriateness, clarity, and alignment with NEC requirements. Revisions were made collaboratively.
4. **Student Trial and Feedback:** Items were trialed with selected students (C1–C2 level) under timed conditions. Post-test feedback was collected to assess task difficulty, engagement, and clarity.

Validation was achieved through expert judgment (via teacher panels), CEFR-level screening, and learner performance data, ensuring that both construct validity and face validity were met.

3.4. Involvement of Teachers and Students

Teachers played a central role as validators, curators, and co-designers. Their tasks included:

- Selecting texts aligned with thematic goals;
- Prompting and editing AI outputs;
- Adjusting task difficulty for different student cohorts;
- Analyzing student responses for diagnostic purposes.

Students contributed through trial participation and feedback provision, helping identify which tasks were engaging, ambiguous, or too difficult. This learner-informed revision loop is central to test validity and pedagogical alignment.

Moreover, this collaborative approach encourages teacher autonomy in AI use and promotes learner agency—key goals in contemporary language education frameworks.

4. AI Integration in Practice

4.1 Pre-processing source materials

In the process of designing high-quality reading and listening tests for advanced learners, pre-processing source materials is a critical step that directly affects the linguistic validity, CEFR alignment, and task quality of the final test items. Several procedures are necessary to prepare raw input for pedagogical use:

- ❖ **Extracting transcripts from video/audio sources lacking built-in captions** is essential for converting spoken content into textual input. Without this step, test developers lack a reliable linguistic reference to analyze lexical range, discourse structure, and potential listening task cues. Manual or AI-assisted transcription ensures accurate representation of spoken data for subsequent editing and adaptation.
- ❖ **Converting PDF, JPG, or printed documents into editable Word files** allows test designers to manipulate and format texts efficiently. Many high-level materials (e.g., academic articles, CPE-level exercises) exist only in static or image-based formats. Using OCR tools to digitize these documents ensures that content can be restructured, annotated, and tailored to specific assessment needs.
- ❖ **Screening C1–C2 lexical items from reading passages or listening transcripts** is crucial for ensuring that the input texts and test items match the target proficiency level. Using tools like the Oxford Text Checker or English Vocabulary Profile enables designers to avoid vocabulary that is too basic or inappropriately advanced, thus preserving level consistency and construct validity.
- ❖ **Screening C1–C2 grammatical items** serves a similar purpose. By identifying advanced grammatical structures (e.g., inversion, nominalisation, reduced relative clauses), test developers can align text complexity with CEFR descriptors. This step also helps in designing grammar-based tasks or syntactic awareness questions that reflect the linguistic demands expected of C1–C2 learners.

Overall, this pre-processing stage ensures that source materials are both technically editable and pedagogically appropriate, laying the foundation for reliable, level-appropriate, and cognitively

engaging test design. Skipping or overlooking these steps can lead to mismatches in task difficulty, affect learner performance validity, and reduce the instructional value of the test.

- Extracting transcripts from video/audio sources lacking built-in captions.
- Converting PDF, JPG, or printed documents into editable Word files
- Screening C1–C2 lexical items from reading passages or listening transcripts based on CEFR levels.
- Screening C1–C2 grammatical items from reading passages or listening transcripts based on CEFR levels.

a. Extracting transcripts from video/audio sources lacking built-in captions.

Instruction 1: How to Use **Sider** to extract transcripts from videos on websites which provide no subtitles

Step 1: Install Sider Extension

- Open Google Chrome.
- Go to Chrome Web Store.
- Search for “**Sider: ChatGPT Sidebar & Copilot**”.
- Click “Add to Chrome” → “Add Extension”.

Step 2: Sign In

- Click the Sider icon in the top-right corner of your browser.
- Sign in using Google or Email.

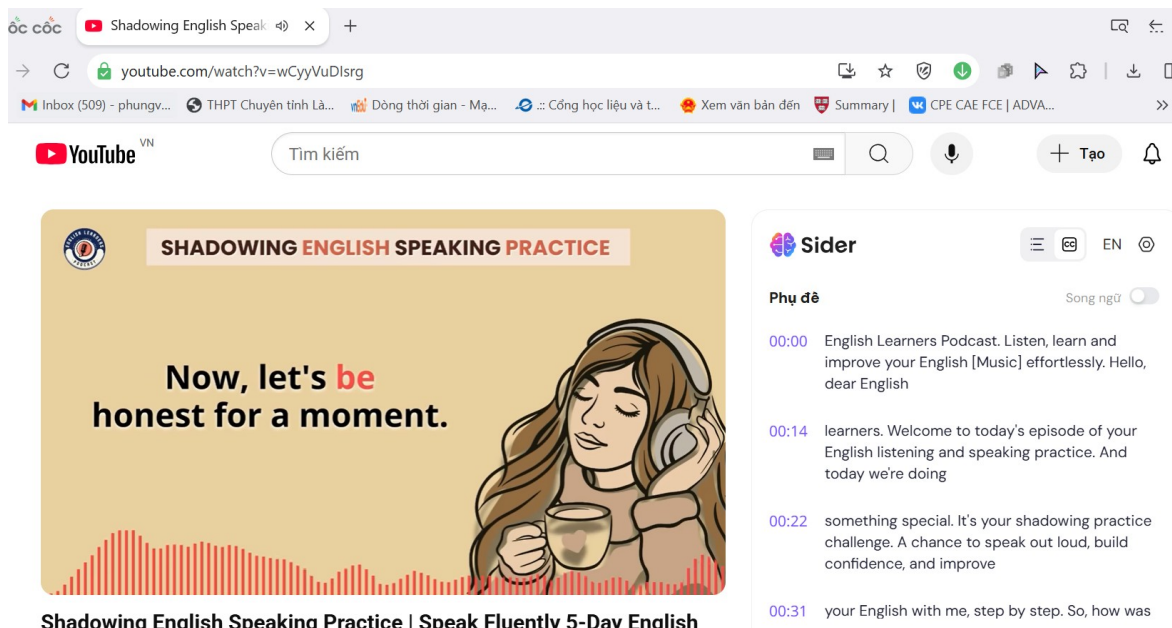
Step 3: Open the Video

- Go to the video page (YouTube, Coursera, Udemy, etc.).
- Play the video.
- Open Sider and choose the “YouTube Summary” or “Video Summary” tab.
- Sider will auto-detect and show the full transcript (if available).

Step 4: Copy or Use the Transcript

- Copy the full transcript.
- Ask Sider to summarize, translate, or analyze the text (If necessary)

Illustration 1: A practical example of extracting transcript from Youtube video with Sider



The screenshot shows a web browser window with a YouTube video player. The video title is "SHADOWING ENGLISH SPEAKING PRACTICE". The video content shows a woman with long brown hair wearing headphones and holding a cup of coffee. The Sider extension sidebar is open on the right, displaying a transcript of the video. The transcript includes the following text:

Phụ đề Song ngữ

00:00 English Learners Podcast. Listen, learn and improve your English [Music] effortlessly. Hello, dear English

00:14 learners. Welcome to today's episode of your English listening and speaking practice. And today we're doing

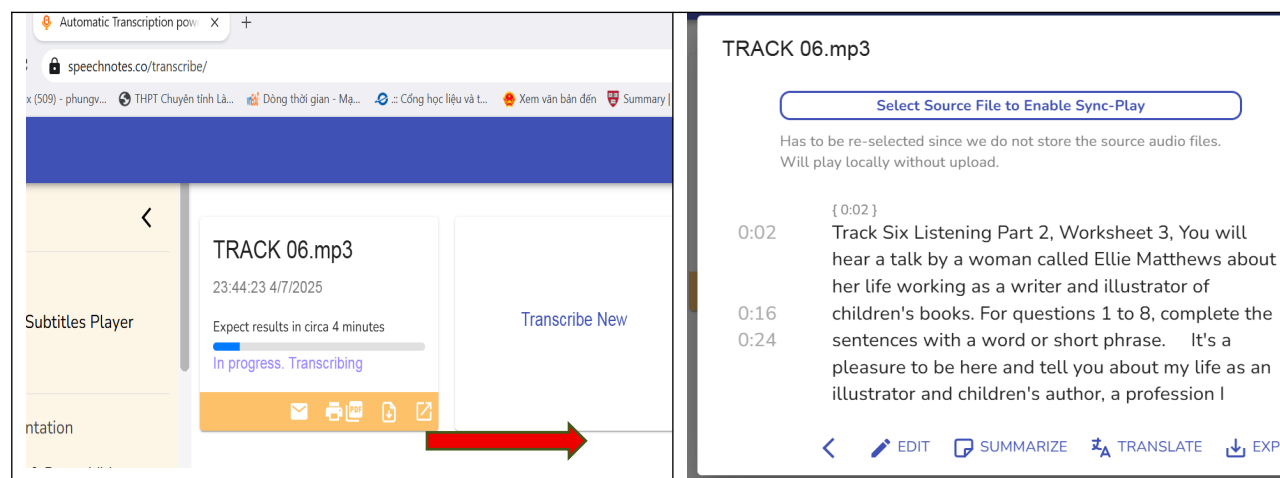
00:22 something special. It's your shadowing practice challenge. A chance to speak out loud, build confidence, and improve

00:31 your English with me, step by step. So, how was

Instruction 2: How to Use **Speechnotes** to extract transcripts from any audio file in English

Step 1: Go to Speechnotes Website 1. Open your browser and go to https://speechnotes.co. 2. Click “Start Dictation” to open the main tool.
Step 2: Set Input Language In the top menu, choose “English” as the input language for accurate transcription.
Step 3: Play Your Recording 1. Open your audio file (on another device or another browser tab). 2. Place your microphone close to the speaker OR use Stereo Mix (if available) to capture internal audio.
Step 4: Let Speechnotes Transcribe As the audio plays, Speechnotes will automatically transcribe the speech into text in real time.
Step 5: Copy or Download the Transcript When finished, you can: ○ Copy the text. ○ Download it as a .txt or .doc file. ○ Email or save to Google Drive directly from the site.

Illustration 2: A practical example of extracting transcript from an uploaded mp3 audio file



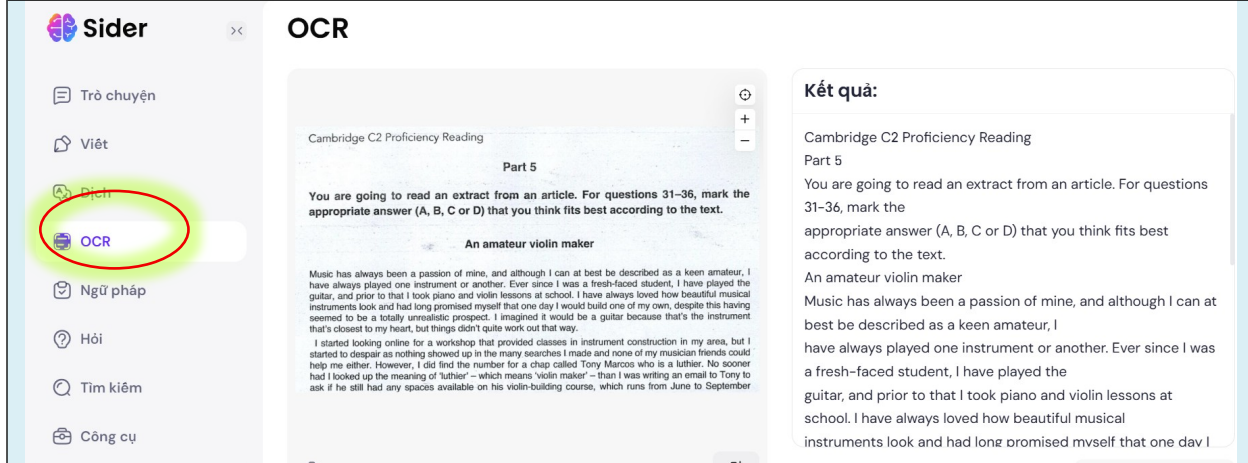
b. Converting PDF, JPG, or printed documents into editable Word files

Instruction 3: Converting PDF, JPG, or printed documents into editable Word files with Sider OCR, ChatGPT and Microsoft Word

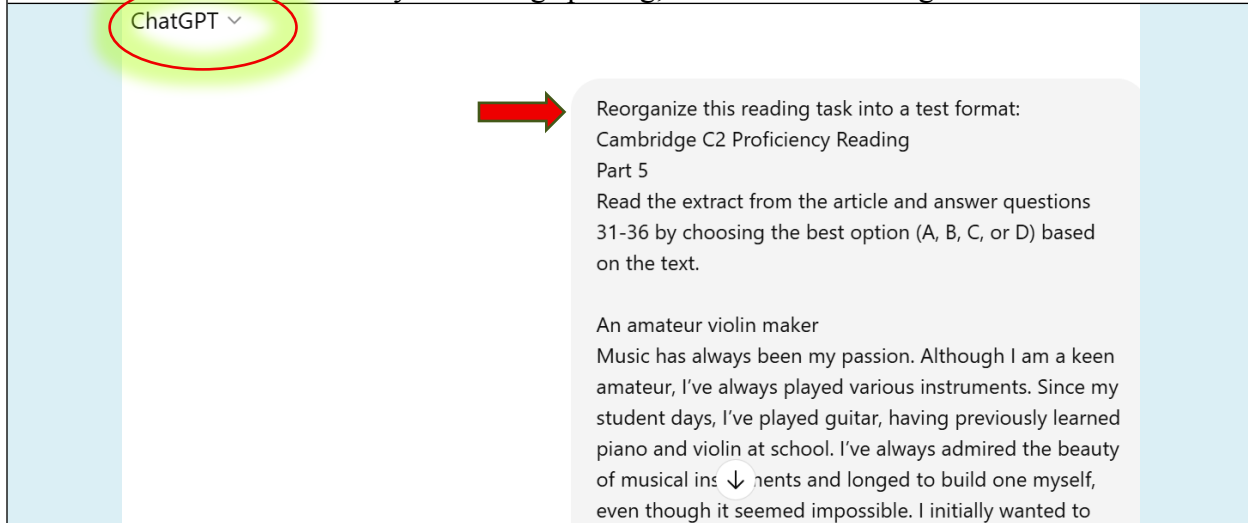
Step 1: Select the PDF, JPG, or printed file you wish to convert.
Step 2: Open Sider (ChatGPT’s sidebar tool introduced in 3.1.1), then upload the file to extract text using OCR .
Step 3: Use ChatGPT to reformat the extracted content into a clean, editable exam paper, including basic layout adjustments and spell check.
Step 4: Copy the final version into Microsoft Word for any final formatting refinements and customization.

Illustration 3: A practical example of creating an editable Word doc for a reading task taken from a printed book

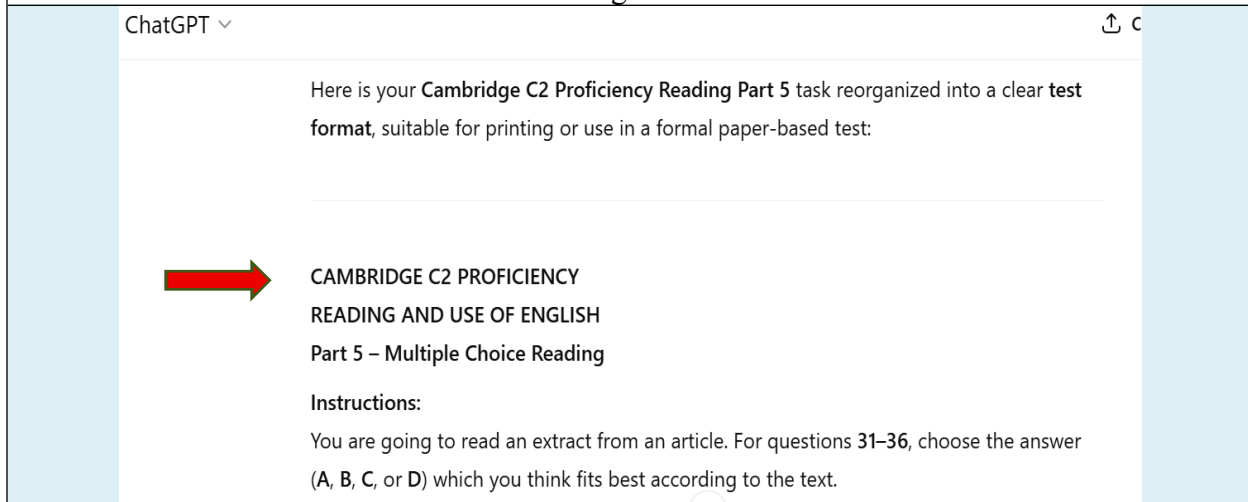
STEPS 1+2: Install Sider AI → upload the image or PDF file
→ extract the Word version using Sider's OCR tool.



STEP 3: Copy the extracted text into ChatGPT and **prompt** it to redesign the test, automatically correcting spelling, and basic formatting errors.



STEP 4: Copy the revised content from ChatGPT into Microsoft Word to finalize the formatting of the test.



c. Screening C1–C2 lexical items from reading passages or listening transcripts based on CEFR levels.

Instruction 4: How to screen C1–C2 lexical items from reading passages or listening transcripts, using Oxford Text Checker and ChatGPT

- Step 1:** Select a reading text you want to analyze for advanced vocabulary.
Step 2: Paste the text into the **Oxford Text Checker** and run the CEFR-level analysis.
Step 3: Filter the results to display only C1 and C2 level vocabulary using the *Activities > Filter > C1, C2* option.
Step 4: Choose *Create a word list* or *Create an exercise* to recycle the selected vocabulary into targeted learning materials.
Step 5: Use **ChatGPT** to design a detailed glossary with phonetic transcription, example sentences, and meanings for each C1–C2 word.

Illustration 4: A practical example of screening C1/C2 vocabulary from a reading text

Step 1: Input the text				Step 2+3: Run Oxford Text Checker
Part 2. Read the following passage and do the tasks that follow. THE REINVENTION OF A BRAND Almost everyone knows what Lego is and has probably played with it at some point in their lives. Since the company's foundation in 1932, millions have been delighted by the interlocking bricks that can unleash the creative ideas lurking inside us, adults and children alike. Part of the attraction is in the attention to detail, which of course the customer is largely unaware of. The Danish creator, Ole Kirk Christiansen, was meticulous about 'doing things right', which is one of the reasons why this popular toy has stood the test of time. However, this hasn't always been the case. Despite its huge success during the 1970s and 1980s, by the beginning of the twenty-first century, the company's fortunes were looking decidedly precarious. It was in debt to the tune of \$800 million, and sales were in fast decline. What had seemed inconceivable throughout the company's history, a total and utter collapse was beginning to look like reality. What led to this unfortunate state was a series of bad decisions based on advice given by external business consultants. At this time, product diversity and business expansion were very much in fashion in all sorts of different industries and so The Lego Group had started to venture into various sectors that lay well outside its expertise. It had created theme parks, and clothes and jewellery for none of which were creating significant revenues so it was clearly time to get back to basics and rethink their product strategy. In order to do this, the newly appointed CEO set about rebuilding the organisation. Business operations were streamlined, and many things were scaled back such as staff and the product ranges. In addition, the company got rid of all the extras that were not a core part of the business, which meant that it was able to return to its roots. This reinvention of the company led to a resurgence in sales and a return to profitability.				Word list: Oxford 5000 CEFR level: A1 A2 B1 B2 C1 THE REINVENTION OF A BRAND Almost everyone knows what Lego is and <u>has</u> probably played with <u>it</u> at some <u>point</u> in the <u>past</u>. <u>Since</u> the company's foundation in 1932, millions <u>have been delighted by</u> the interlocking <u>bricks</u> that <u>can</u> unleash the <u>creative ideas</u> lurking <u>inside us</u>, adults and children <u>alike</u>. Part of the <u>attraction</u> is in the <u>attention to detail</u>, which of course the customer is largely unaware of. The <u>Danish creator</u>, Ole Kirk Christiansen, was meticulous about 'doing things right', which is one of the <u>reasons why</u> <u>this popular toy</u> <u>has</u> stood the test of <u>time</u>. However, <u>this hasn't</u> always been the case. Despite its huge success during the 1970s and <u>by</u> the beginning of the twenty-first century, the company's fortunes were looking decidedly precarious. <u>It</u> was in debt to the tune of \$800 million, and <u>sales were</u> in fast decline. What <u>had seemed inconceivable throughout</u> the company's history, a <u>total</u> and utter collapse was beginning to <u>look like</u> reality. What led to <u>this unfortunate state</u> was a series of bad <u>decisions</u> based on advice given <u>by</u> external business consultants. At <u>this time</u>, product diversity and business expansion were <u>very much</u> in fashion in <u>all sorts</u> of different industries and so <u>The Lego Group</u> had started to venture into various sectors that lay well outside its expertise. It had created theme parks, and clothes and jewellery for none of which were creating significant revenues so it was clearly time to get back to basics and rethink their product strategy. In order to do this, the newly appointed CEO set about rebuilding the organisation. Business operations were streamlined, and many things were scaled back such as staff and the product ranges. In addition, the company got rid of all the extras that were not a core part of the business, which meant that it was able to return to its roots. This reinvention of the company led to a resurgence in sales and a return to profitability.
Step 4+5: filter C1/C2 word list → copy to and prompt ChatGPT to create a glossary				
Word	CEFR Level	/Phonetic/	POS	Example Sentence
alike	C1	/ ə 'laɪk /	adv	Good management benefits employers and employees <i>alike</i> .
alike	C2	/ ə 'laɪk /	idiom	A few months ago, I was working on a group presentation with a classmate. As we started brainstorming, we both suggested almost the exact same opening slide at the same time. I laughed and said, ' <i>Great minds think alike!</i> ' I guess it showed how well we understood the topic and how similarly we approached problems.
creator	C1	/ /kri 'eɪ.tər/	noun	Walt Disney, the <i>creator</i> of Mickey Mouse

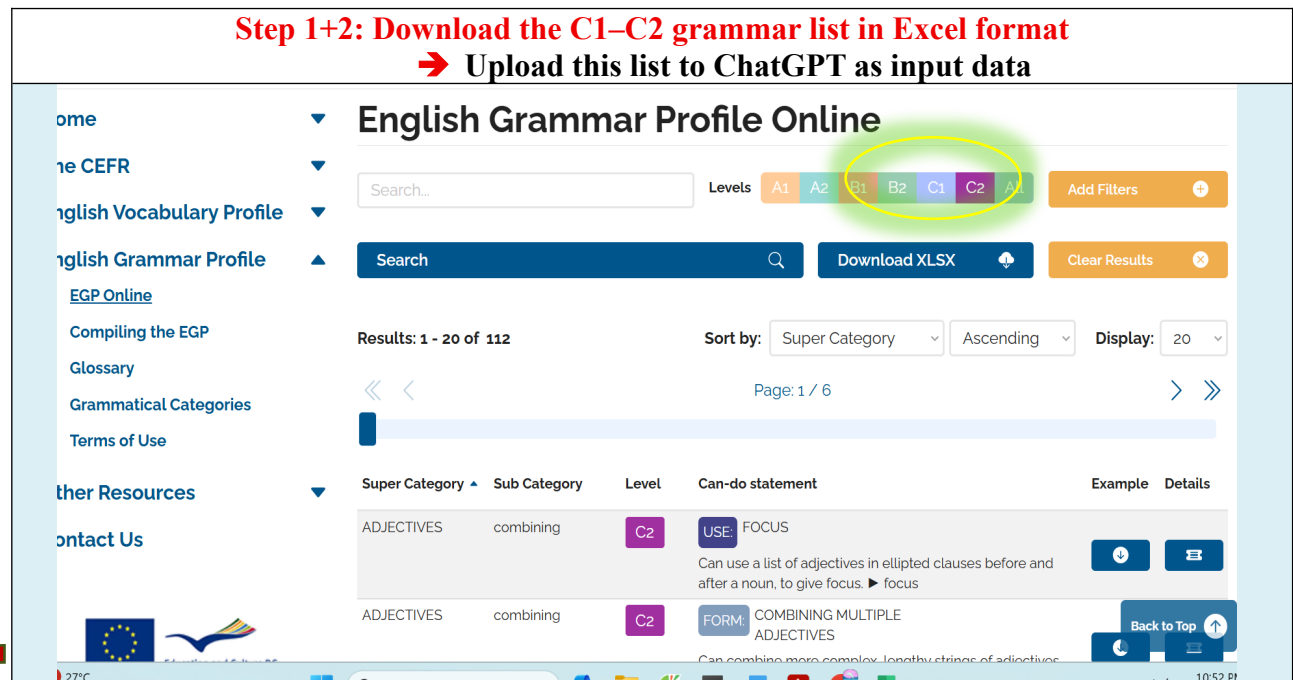
d. Screening C1–C2 grammatical items from reading passages or listening transcripts based on CEFR levels.

Instruction 5: How to screen C1–C2 grammatical items from reading passages or listening transcripts, using English Profile and ChatGPT

Step 1: Download the C1–C2 grammar list in Excel format from englishprofile.org.
Step 2: Upload the Excel file to ChatGPT as input data.
Step 3: Prompt ChatGPT to use the uploaded grammar list to identify and extract all C1–C2 grammar items from your target text.

Illustration 5: A practical example of prompting for screening C1-C2 grammar from a text

Step 1+2: Download the C1–C2 grammar list in Excel format
➔ Upload this list to ChatGPT as input data



The screenshot shows the 'English Grammar Profile Online' interface. The 'Levels' filter is set to 'C1' and 'C2'. The 'Download XLSX' button is highlighted. Below the interface, an Excel table provides the grammar list data.

Super Category	SubCategory	Level	Guideword	Can-do statement	Example
ADJ	combining	C2	USE: FOCUS	Can use a list of adjectives in elliptical clauses before and after a noun, to give focus. ► focus	<i>Imaginative, good-humoured, friendly, he is liked by his peer-students. (Poland; C2 MA Polish; Pass)</i>
ADJ	combining	C2	FORM: COMBINING MULTIPLE ADJECTIVES	Can combine more complex, lengthy	<i>A timid, shy, self-conscious, over-sensitive</i>
ADJ	comparatives	C2	FORM/USE: WITH 'NO' OR 'NOT ANY'	Can use 'no' / 'not any' with comparati	<i>That is why we are not any happier than yo</i>
ADJ	comparatives	C2	FORM/USE: WITH 'NOT THAT MUCH'	Can use 'not that much' to modify con	<i>Although she is not that much older than I</i>
ADJ	modifying	C2	FORM: COMPARISON	Can use adjectives in 'as ... as' and 'so	<i>Because our debates do not draw as large</i>
ADJ	position	C2	FORM/USE: DEGREE ADJECTIVES BEFORE	Can use an increasing range of adjecti	<i>As a conclusion, I shall say that the major</i>
ADJ	superlatives	C2	FORM/USE: 'SLIGHTEST', 'FAINTEST'	Can use 'the slightest', 'the faintest' wi	<i>I should not end my relationship with Alex</i>
ADV	adverbs and adve	C2	USE: DISTANCING	Can use adverbs mid clause, to distan	<i>He came out from school at the age of 16, e</i>
ADV	position	C2	FORM/USE: MID POSITION, DISTANCING	Can use adverbs in mid position, to di	<i>In all actuality Piggy himself theoretically</i>
ADV	position	C2	FORM/USE: FRONT POSITION, INVERSION, V	Can use 'hardly' in front position follo	<i>Hardly had he reached his front door when</i>
ADV	adverb phrases -	C2	FORM: PRE- AND POST-MODIFIED ADVERB	Can form adverb phrases with modifie	<i>Many politicians and owners of big enterp</i>

Step 3: Paste a text → prompt ChatGPT to analyse C1/C2 grammar items

Part 5. For questions 86 – 95, read the following passage and choose from the sections (A – E). The sections may be selected more than once. Write the letter A, B, C, D, or E in the corresponding numbered boxes provided.

ADDRESSING THE PROBLEM OF WATER SCARCITY

A. Despite 70% of the world's surface being covered with water, less than 3% of that is actually drinkable and, based on factors ranging from agricultural practices to climate change and daily habits, water scarcity is fast becoming a serious problem in many countries around the globe. The problem of scarcity – in other words, having insufficient water – is categorised as being physical (happening in places where supply cannot meet demand) or economic (occurring in areas that have plenty of water but don't have good management systems). Broadly speaking, the causes of water scarcity are related to the rapid rise of the global population and the associated issues that this has brought, and the predictions in many parts of the world are that the population will continue to rise for several decades. This suggests that rather than lamenting the journey to this point, we would be better off focusing on solutions, especially given that the causes have been written about extensively and are very well understood, which is not the case for the solutions.

B. Solutions for water scarcity should primarily be context-specific if they are to work, and must include experts, organisations and charities that can provide insight into the particular challenges. For example, there is no point in poorer nations engaging expert engineers from overseas to help with the infrastructure for water management systems if the resulting system is not affordable or able to withstand the climatic conditions of the region. Too often collaborations on projects like this turn into vanity projects for the foreign companies in much the same way as high-end commercial projects, such as the building of a luxury hotel or impressive bridge. Workable solutions will need to understand the influences of geology, the environment and the influences of local culture to be successful and to be a combination of technology and changes in human behaviour. An immediate approach is to provide

ChatGPT

C1–C2 Grammar Structures Identified

Grammar Item	Example from Text	CEFR Level	Explanation
Complex passive voice	"...is categorised as being physical..."	C1	Use of passive structure with "as being" + adjective for abstract classification.
Reduced relative clause	"...challenges that this has brought..."	C1	Use of defining relative clause embedded in a complex noun phrase.
Cleft structure	"What is not the use are the solutions." (implied: "...which	C2	Cleft or pseudo-cleft emphasis on contrast between known and

4.2. A step-by-step practical example of recycling a CPE Reading task into a Summary Completion Listening task using AI tools

Before designing a listening task aligned with actual national-level exams, the author analyzed the English test from the 2024–2025 National English Contest (NEC) to identify the number of questions, types, formats, and target skills. Based on the statistical data obtained, teachers can formulate specific prompts for AI tools to assist effectively in designing question patterns and uncovering the underlying purposes and implications behind each question type. Whether constructing listening or reading items, the design approach remains fundamentally similar.

Illustration 6: Analysing Listening Question Types in NEC 2024 - 2025

Skill	Question Type	Response Format	Cognitive Skill	Number of Questions
Listening	Speaker Identification	Letter (M/F/B)	Viewpoint analysis	5

Listening	Matching Information	Matching A–J	Inference & association	5
Listening	Multiple Choice	A/B/C/D	Gist, detail, inference	5
Listening	Summary Completion (Gap-fill)	Max 3 words	Lexical precision, paraphrasing	10

The selected source text is originally a dense argumentative article titled *"Addressing the Problem of Water Scarcity"*, published in **Prosperity Education's** CPE preparation coursebook. The author recycled this passage into a Listening Summary Completion task with the following steps. The original reading passage is included in the Appendices.

Step 1: The text was first pre-processed using Sider/OCR to extract editable text from the printed page. Teachers simply need to paste the image to Sider, and then prompt: OCR.

The screenshot shows the Sider Fusion interface. On the left, a document titled "Addressing the Problem of Water Scarcity" is displayed. A red arrow points from the document to the OCR output on the right. The OCR output is a transcribed version of the document's text, starting with "Despite 70% of the world's surface being covered with water, less than 3% of that is actually drinkable..." and ending with "...the causes of water scarcity are relat".

Step 2: Recycling the transcript targeted for listening text with Sider

Key sections were then reformulated into a natural spoken script, with sentence structures and vocabulary adapted to suit listening comprehension rather than silent reading, still with Sider. Command prompt used: *"Please reformulate the above text into a natural spoken script, with sentence structures and vocabulary adapted to suit listening comprehension rather than silent reading. The length of the transcript is suitable for a 3-minute recording. Summarize or develop the script to meet the length if needed"*

Here is a part of the transcript generated by Sider, the full version included in the Appendices.

The screenshot shows the Sider interface with a reformulated transcript for "Addressing Water Scarcity". The transcript is a natural spoken script, starting with "Did you know that although about 70% of the Earth's surface is covered with water, less than 3% of that is actually drinkable? And even then, many countries are facing serious water shortages. Why? Well, water scarcity isn't just about not having enough water — it's also about how we manage it. Sometimes, there's plenty of water, but it's in the wrong place, or we can't get to it. This can happen because of climate change, growing populations, or poor management practices."

Most solutions need to be specific to each area. We can't just copy and paste a plan from one country to another. For example, if a community doesn't have reliable infrastructure or good technology, then fancy projects won't work. We also need to understand local culture and the environment. Only then can we develop effective solutions....

Step 3: Screen C1-C2 grammar content in the transcript with ChatGPT (prompted with English Profile C1-C2 grammar table input in advance)

ChatGPT

Here's a C1-C2 grammar analysis of the transcript titled "Addressing Water Scarcity: A Conversation." The focus is on identifying advanced grammatical structures typically aligned with CEFR C1-C2 descriptors, particularly those noted in the English Profile (<https://www.englishprofile.org>).

C1 & C2 Grammar Features Identified

Sentence / Excerpt	CEFR Level	Grammar Feature	Notes
"Although about 70% of the Earth's surface is covered with water, less than 3% of that is actually drinkable."	C1	Concessive clause with "although"	Clear use of contrastive subordination, typical at upper-intermediate/advanced level.

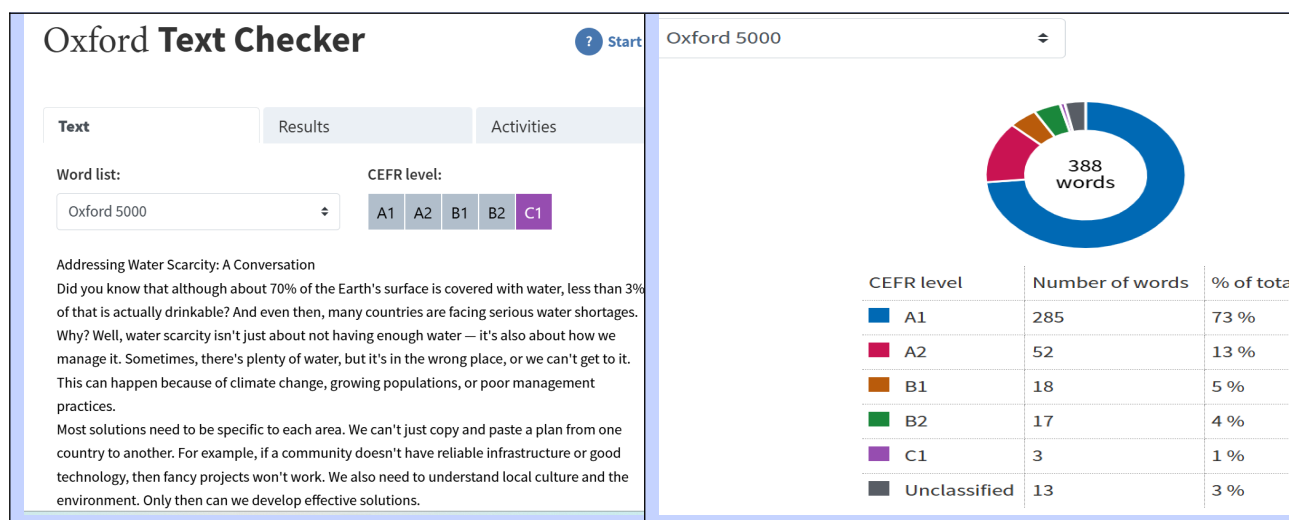
The analysis provided by ChatGPT regarding grammar showed positive feature with multiple advanced grammar items detected from the transcript, *fully listed by ChatGPT in the table below*:

Sentence / Excerpt	CEFR Level	Grammar Feature	Notes
"Although about 70% of the Earth's surface is covered with water, less than 3% of that is actually drinkable."	C1	Concessive clause with "although"	Clear use of contrastive subordination, typical at upper-intermediate/advanced level.
"Well, water scarcity isn't just about not having enough water — it's also about how we manage it."	C1	Cleft-like emphatic structure + verb pattern "be about + -ing"	The dual-focus expression and verb pattern are characteristic of nuanced explanation.
"Sometimes, there's plenty of water, but it's in the wrong place, or we can't get to it."	C1	Contrast and ellipsis	Coordination with omitted repeated structures, managing complexity.
"We can't just copy and paste a plan from one country to another."	C1	Modal + bare infinitive (can't just copy)	Informal register but controlled modal usage with nuanced implication.
"Only then can we develop effective solutions."	C2	Inversion after fronted adverbial	High-level inversion used for emphasis — typical C2 feature.
"If a community doesn't have reliable infrastructure or good technology, then fancy projects won't work."	C1	Conditional sentence (first conditional)	While structurally simple, conditional logic tied to complex ideas is a C1 marker.
"Installing water meters can give families an incentive to save water, because the cost is directly linked to how much they use."	C1	Nominalization of -ing clause + passive voice	Combining multiple advanced elements in one sentence: "installing..." (subject), and "linked" (passive).
"Many people don't realize how much water goes into producing the	C1	Embedded WH-clause ("how much water goes	Requires parsing nested clause + verb-noun collocation.

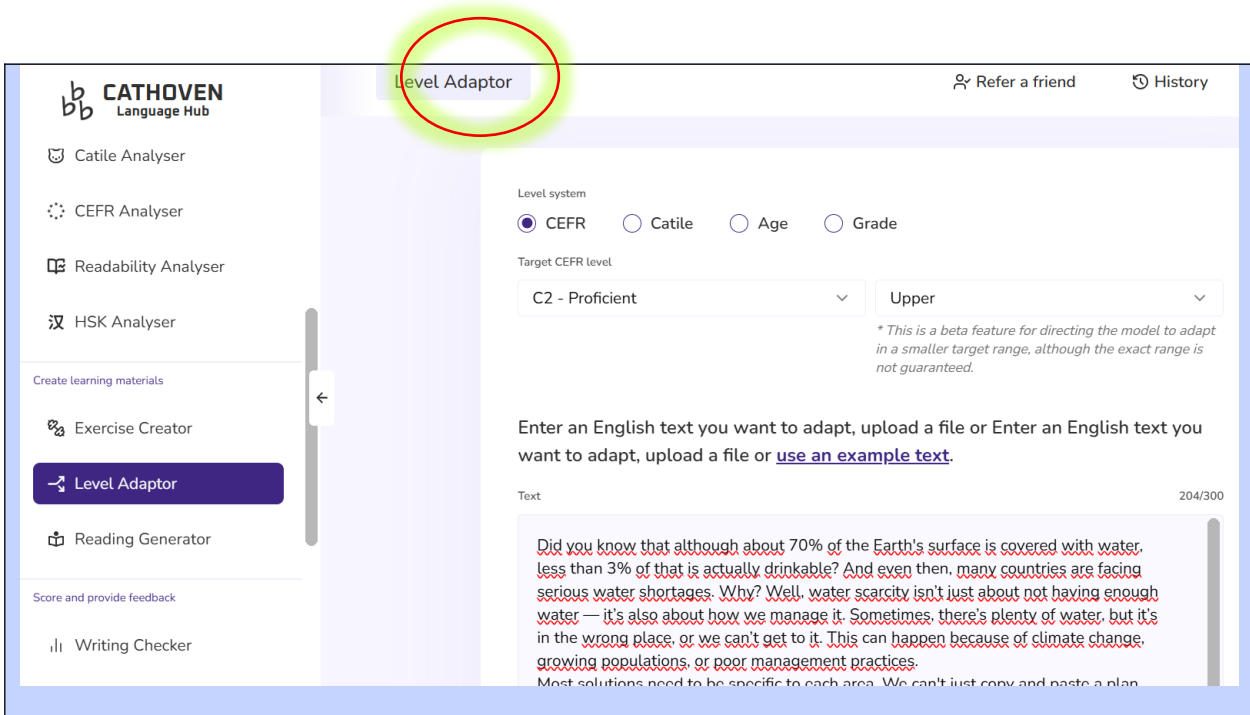
food they eat.”		into producing...”)	
“Simple changes, like collecting rainwater..., can make a big difference.”	C1	Reduced relative clause (“like collecting rainwater”)	Use of gerund as noun modifier and implication without explicit relative pronouns.
“Protecting these areas is crucial.”	C1	Gerund as subject	Use of non-finite clause as subject shows flexibility with noun phrases.
“At the end of the day, solving water scarcity is about combining smart technology...”	C1	Gerund phrase as subject complement	Parallel -ing structure listing actions: combining, educating, protecting.
“If we all work together, we can make a real difference.”	C1	First conditional again, but used in a motivational tone	Common in speech but used here to project hypothetical consequence.

Step 4: Screen the C1-C2 vocabulary with Oxford Text Checker

Then the author used Oxford Text Checker to assess the lexical items in the transcript and the result was that advanced vocabulary (C1) consisted only 1% of the transcript.



While the grammatical structures in the transcript met the required standards, the lexical range was deemed insufficient. Consequently, the author employed **Cathoven AI/Level Adaptor** to enhance the vocabulary



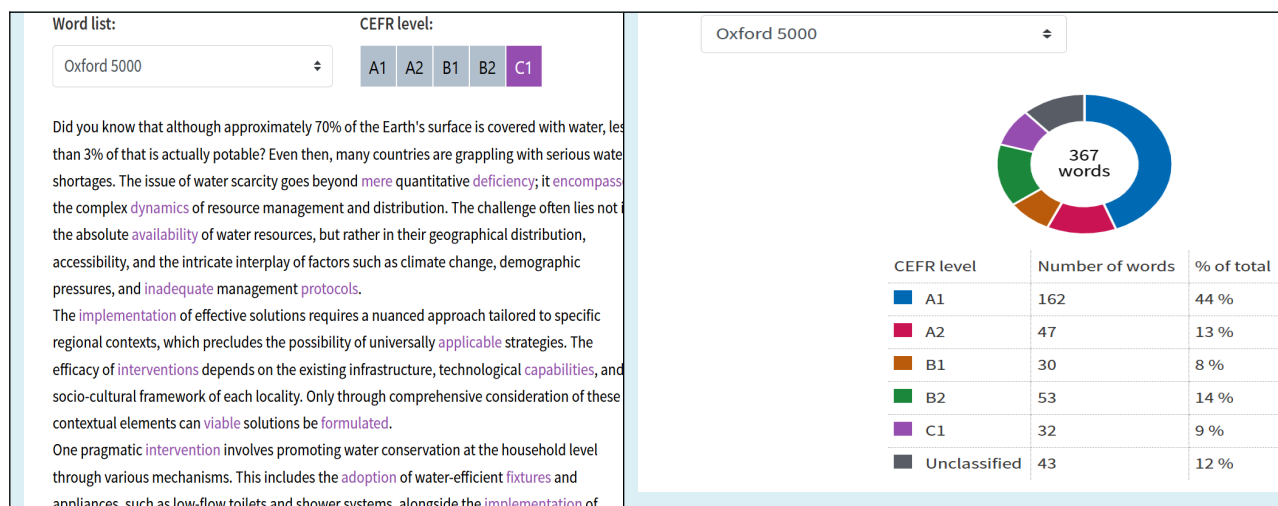
Here is a part of the final transcript generated by **Cathoveen/Level Adaptor**, the full version included in the Appendices.

Addressing Water Scarcity

Did you know that although approximately 70% of the Earth's surface is covered with water, less than 3% of that is actually potable? Even then, many countries are grappling with serious water shortages. The issue of water scarcity goes beyond mere quantitative deficiency; it encompasses the complex dynamics of resource management and distribution. The challenge often lies not in the absolute availability of water resources, but rather in their geographical distribution, accessibility, and the intricate interplay of factors such as climate change, demographic pressures, and inadequate management protocols.

The implementation of effective solutions requires a nuanced approach tailored to specific regional contexts, which precludes the possibility of universally applicable strategies. The efficacy of interventions depends on the existing infrastructure, technological capabilities, and socio-cultural framework of each locality. Only through comprehensive consideration of these contextual elements can viable solutions be formulated...

When **double-checked with Oxford Text Checker**, the author satisfied with the availability of 32 advanced vocabulary (9%) in the newly-generated transcript. The transcript was then ready for the next stage of question design.



Step 5: Creating the audio file with TTSMaker

On the next stage, the audio version was generated using AI voice synthesis tools to simulate authentic lecture delivery. The teacher simply pasted the transcript onto TTSMaker <https://ttsmaker.com/>, selected a formal voice and wait for a quick conversion. The recording of 2:52 second length was produced as expected, and could be downloaded in mp3 format.

TTSMaker Free Text to Speech Tiếng Việt Pro Upgrade

Maximum characters 1000 remaining 0 available (Exceeded the maximum limit of 1909 characters) Upgrade to Pro for support of over 20,000 characters per conversion

Clear Text

Did you know that although approximately 70% of the Earth's surface is covered with water, less than 3% of that is actually potable? Even then, many countries are grappling with serious water shortages. The issue of water scarcity goes beyond mere quantitative deficiency; it encompasses the complex dynamics of resource management and distribution. The challenge often lies not in the absolute availability of water resources, but rather in their geographical distribution, accessibility, and the intricate interplay of factors such as climate change, demographic

Language: English

Voices Selected: 148 Try new AI Voice Dialogue Generator

Voice (Slow + Unlimited)

Example Limit Per Conversion: 1000 Characters

3 Emotion 1480 - Alayna-usUnited States Female (V2)

Multi-Emotion Version

Example Limit Per Conversion: 3000 Characters

TTSMaker Free Text to Speech Tiếng Việt Pro Upgrade

convert text to speech success!

0:07 / 2:52

This audio file will be automatically deleted within 30 minutes, please download it in time.

Language: English

Voices Selected: 2599 Try new AI Voice Dialogue Generator

Limit Per Conversion: 3000 Characters

13 2708 - Mason-usUnited States Male

Example Limit Per Conversion: 2000 Characters

14 2598 - Ashe-usUnited States Female

Example Limit Per Conversion: 1000 Characters

Step 6: Creating the questions with ChatGPT

After having the audio file with suitable length, and a transcript of expected CEFR lexico-grammar level, the author came to the final stage of creating the listening paper. To do this, the author pasted the transcript onto ChatGPT and prompted the platform to generate a listening – Summary completion (Gap-fill) with 10 gaps for max 3-word answers, examining lexical precision and paraphrasing.

The output listening test is as below:

Listening Section: Summary Completion (10 questions)

Complete the summary below using NO MORE THAN THREE WORDS from the recording.

Addressing Water Scarcity

Water scarcity is not only a result of limited supply but also of ineffective (1) _____ and distribution systems. In some cases, the problem is not the quantity of water, but where it is and how easily it can be accessed. Influencing factors include (2) _____, population growth, and weak management.

Because each region is different, a (3) _____ is essential. A solution that works in one country may not suit another due to differences in infrastructure, technology, or (4) _____ conditions.

One proposed strategy is to encourage households to use water responsibly. Installing (5) _____ and water-efficient appliances helps reduce usage. (6) _____ schemes can also motivate families to conserve water, linking usage directly to cost. Unfortunately, many governments don't provide enough support for (7) _____ groups.

Awareness campaigns are also vital. People often don't realize how much water is needed to produce their food, especially when it comes to (8) _____. Better education could lead to more informed consumer behavior.

On a small scale, simple practices like (9) _____ can make a significant difference. Moreover, protecting (10) _____, which naturally filter and store water, is crucial for long-term sustainability.



Please scan this QRCode for the audio file.

This AI-generated listening material was subsequently piloted with advanced-level students in a classroom setting, as detailed in Section 5. Case Studies and Sample Tests.

4.3. A step-by-step practical example of recycling an article from online sources into a C1-C2 Reading Cloze Test using AI tools

Similarly, before designing a reading task aligned with actual national-level exams, the author analyzed the English test from the 2024–2025 National English Contest (NEC) to identify the number of questions, types, formats, and target skills.

Illustration 7: Analysing Reading Question Types in NEC 2024 - 2025

Skill	Question Type	Response Format	Cognitive Skill	Number of Questions
Reading	Cloze Test / Word Formation	Fill-in (no choices)	Grammar/Vocab accuracy	10
Reading	True / False / Not Given	T / F / NG	Fact vs. opinion	6
Reading	Summary Completion (open cloze)	1–3 words	Paraphrase, scanning	8
Reading	Sentence Insertion	A–E	Cohesion & logical flow	6
Reading	Matching (People/Ideas)	A–F	Attribution	10

Step 1: The author prompted ChatGPT to list out trusted sources for reading material design for C1-C2 language learners. Below is a categorized list of trusted sources where teachers can extract high-quality texts for reading test design:

a. News & Commentary (Expository, Argumentative, Editorial)

Suitable for inference, tone, opinion vs. fact, and paraphrasing tasks.

Source	Suitable Genres	Notes
The Guardian – Long Reads	Essays, reports, opinion	Deep analysis, ideal for multiple-choice & summary tasks
The Atlantic	Commentary, feature articles	Strong C2-level vocabulary and structure
The Conversation	Academic-based journalism	Written by scholars, good for critical reading
BBC Future / BBC Culture	Pop science, culture, tech	Engaging, varied sentence types

b. Academic & Popular Science

Suitable for reading-to-learn, technical vocabulary, and cause-effect relationships.

Source	Type	Notes
National Geographic	Human geography, science	Visual + text-based prompts
Scientific American	Popular science	Dense texts, good for cloze/sentence insertion
Aeon Essays	Philosophy, science, sociology	Argument-heavy, suitable for high-level inference
Nature News	Research summaries	Advanced but readable for top students

c. Literature & Narrative Non-fiction

Suitable for author's purpose, character inference, and figurative language.

Source	Genre	Notes
The New Yorker – Fiction & Essays	Literary fiction, personal essays	Excellent for tone, voice, style
Granta Magazine	Short stories, memoirs	High literary value

Narratively	Narrative journalism	Story-based, with moral or social angles
Oxford World's Classics (public domain excerpts)	Classic literature	Good for 19th-century vocabulary and sentence structure

d. Education, History, and Society (Cross-curricular themes)

Suitable for test formats aligned with thematic units or Content and Language Integrated Learning (CLIL) approaches.

Source	Subject	Notes
History Today	History essays	Causal reasoning, comparison
Stanford Encyclopedia of Philosophy	Philosophy entries	Use for adapted paraphrase/inference tasks
Edutopia	Education articles	Accessible but still challenging
Foreign Affairs	Politics, international relations	Bias, author's stance, lexical complexity

In this step, the author selected an article titled *"Massive Study Flips Our Story of Addiction and the Brain"* from <https://www.scientificamerican.com/> to create a Reading Cloze Test (open cloze).

Step 2: The author prompted ChatGPT to solve **questions 46 – 55 in NEC 2024 – 2025** to analyze the format and skills assessed in this task.

II.2. READING COMPREHENSION (5.0 points)

Part 1. For questions 46 – 55, read the following passage and fill in each of the numbered spaces with ONE suitable word. Write your answers in the corresponding numbered boxes provided.

Smiles make us smile. They are (46) _____: a smile can be passed from one person to another. We are certain we understand what a smile means. It's an (47) _____ of happiness or pleasure, easy to spot and easy to understand, and it charms us. People seem at their (48) _____ natural and spontaneous when they smile – we all know what a forced smile looks like. We smile automatically when someone points a camera at us; it's a social and cultural reflex, and what we (49) _____ of a picture portrait. Yet the smile, like everything else, has a history. If you walk around any art gallery and look at the portraits, you'll find very (50) _____ genuine smiles. The same is true of early photograph portraits. They were, it seems, unfashionable for a long period in the history of portraiture.

Physically, a smile (51) _____ is not complex – it comes when we contract the *zygomatic major*, a muscle in the face – but the meaning of smiles changes with the mood of the times, and can even be different for men and women in the (52) _____ culture. Still walking around the gallery, you will find more smirks than smiles. One reason for this may be that with a smile, what you see is what you get, a frozen moment of happy unselfconsciousness, and (53) _____ a smirk allows for several interpretations – condescension, flirtatiousness, boredom, and so on – in other words, it is more (54) _____ and forces a longer and deeper engagement with the painting from the (55) _____.

And here is the question analysis by ChatGPT:

No	Word	Skill/Knowledge Tested
46	<i>infectious</i>	Collocation awareness & metaphorical meaning – understanding that “infectious” can describe emotions or expressions like a smile, not just diseases.
47	<i>expression</i>	Lexical accuracy & noun collocation – selecting the correct abstract noun that fits the phrase “an expression of happiness.”
48	<i>most</i>	Grammar: intensifiers – mastery of fixed structures like “at their most + adjective” to express extremes.
49	<i>expect</i>	Verb + preposition use – proper usage of “expect of someone/something,” common in evaluative or descriptive contexts.

50	<i>few</i>	Quantifier accuracy – correct selection of a determiner that logically matches the context (“genuine smiles”) and negative frequency.
51	<i>itself</i>	Emphatic/reflexive usage – understanding how “itself” adds contrast or emphasis (C1 grammar awareness).
52	<i>same</i>	Comparative reference – recognizing contrast within the same culture vs. cross-cultural variation (discourse cohesion).
53	<i>whereas</i>	Linking and contrast – appropriate use of a subordinating conjunction to signal a contrast between smile and smirk.
54	<i>ambiguous</i>	Advanced vocabulary & semantic nuance – knowledge of abstract adjectives describing interpretative complexity .
55	<i>viewer</i>	Precise vocabulary – selecting the correct noun for a person interacting with visual art; vocabulary linked to arts and perception.

Step 3: The author pasted the criteria in the column “Skill/Knowledge Tested” above, together with the selected reading passage in step 1, and prompt ChatGPT to create a similar reading task with the questions of the same criteria.

Here is the Reading task generated by ChatGPT:

READING COMPREHENSION

Part 1. For questions 46–55, read the following passage and fill in each of the numbered spaces with ONE suitable word.

For years, the belief that early drug use harms the brain has been deeply (46) _____ - a narrative reinforced by studies linking adolescent substance use to long-term damage. Yet a new study from the U.S. may flip that assumption. Researchers found brain differences existed even before children took their first puff or sip—suggesting certain traits may be (47) _____, not caused by drug use. These findings are (48) _____: they offer a new way of looking at addiction—not as an inevitable outcome of drug exposure, but as something shaped by biology and personality. One significant discovery was that teens who later used drugs had more folded cortices and larger brain volumes—features often (49) _____ with intelligence, risk-taking, and curiosity. Children appeared to be at their (50) _____ vulnerable when openness to experience combined with impulsivity—a mix that increased their likelihood of early experimentation. While the study itself stops short of proving causality, researchers do (51) _____ more flexible prevention strategies focused on high-risk groups. A separate school-based program showed promise. Students were tested for traits like anxiety sensitivity or sensation-seeking, and workshops were delivered months later with no direct (52) _____ to the test. The results: an 87% drop in the (53) _____ of substance use disorder over five years. Researchers stress that traits like curiosity or boldness shouldn’t be (54) _____; these same features drive discovery and innovation. The challenge lies in teaching young people to channel them—without labelling the traits as problematic (55) _____ themselves.

Step 4: The author prompted ChatGPT to provide all plausible answers, together with explanations for the above task:

All possible answers for the reading task provided by ChatGPT:

Question	Best Answer	Other Acceptable Alternatives	Explanation
46	<i>ingrained</i>	<i>embedded, entrenched, accepted, widespread</i>	Collocation: “deeply ingrained belief” is most natural; others express durability or social acceptance.
47	<i>preexisting</i>	<i>inherent, innate, inborn, genetic</i>	All imply that the traits existed before drug use began.
48	<i>infectious</i>	<i>compelling, powerful, transformative, enlightening</i>	"Infectious" metaphorically fits the idea of spreading new ideas; others imply emotional or intellectual impact.
49	<i>associated</i>	<i>linked, connected, correlated, tied</i>	Standard collocations: “associated with intelligence,” etc.
50	<i>most</i>	<i>particularly, especially, extremely (less precise)</i>	“At their most vulnerable” is a fixed C1 structure. Others weaken grammatical precision but retain meaning.
51	<i>recommend</i>	<i>propose, advocate, support, suggest</i>	All are valid verbs of recommendation in academic contexts.
52	<i>reference</i>	<i>mention, link, relation, connection</i>	“Reference to the test” is the most idiomatic; others are acceptable with a slight shift in nuance.
53	<i>odds</i>	<i>likelihood, rate, risk, chance(s), incidence</i>	“Drop in the odds” is idiomatic; others are statistically acceptable and contextually logical.
54	<i>stigmatized</i>	<i>demonized, criticized, pathologized, shamed</i>	All reflect negative social judgment of traits.
55	<i>in</i>	<i>./.</i>	“In themselves” is a C1 academic phrase, focuses on essential nature.

After numerous attempts to leverage AI tools in designing reading tests, the author has come to realize that creating high-quality reading assessments is significantly more complex and challenging to evaluate than applying AI to listening material development. Unlike listening tasks—where input is typically linear, temporally bound, and often based on surface-level comprehension—reading tests demand a more intricate balance of cognitive load, lexical density, question variety, and coherence between input and output. Ensuring that reading tasks align with CEFR descriptors, maintain lexical and grammatical progression, and assess multiple subskills (such as inference, cohesion, paraphrasing, and vocabulary sensitivity) requires far more nuanced calibration. Furthermore, evaluating the quality and level-appropriateness of reading questions—especially open cloze, sentence insertion, or summary completion—poses additional difficulties that AI still struggles to fully resolve without expert human oversight.

5. Case Studies and Sample Tests

This section presents a practical example of how a high-level reading passage was transformed into an advanced listening assessment using AI tools and CEFR-aligned screening. The case study demonstrates how test designers can repurpose existing academic materials to generate listening input that aligns with the cognitive and linguistic demands of C1–C2 learners. Due to the limit of the paper, the author decided to explain only one case study for listening test using recycled materials with AI supports.

5.1. Sample Listening Task: Recycled from Reading

Source Text: “Addressing the Problem of Water Scarcity” (CPE Prosperity Education, p.52)

Original Format: Academic reading passage (~700 words), featuring complex arguments on global water challenges, policy shortcomings, and sustainable practices.

Transformed Format: Listening Summary Completion – 10 gaps

Transformation Process: The source reading was first processed using OCR (Sider) and then fed into ChatGPT with prompts to convert it into a spoken-style summary, preserving main arguments and terminology while simplifying sentence structure for oral delivery. The transcript was then synthesized into audio using TTSMaker, producing a neutral-accent academic tone at ~140 words per minute.

5.2. Sample Task and Answer Key

Task Type: Summary Completion

CEFR Level: C1–C2

Skills Assessed: Global listening, lexical recognition, paraphrase matching, and high-level information processing

Listening Section: Summary Completion (10 questions)

Complete the summary below using NO MORE THAN THREE WORDS from the recording.

ADDRESSING WATER SCARCITY

Water scarcity is not only a result of limited supply but also of ineffective (1) _____ and distribution systems. In some cases, the problem is not the quantity of water, but where it is and how easily it can be accessed. Influencing factors include (2) _____, population growth, and weak management.

Because each region is different, a (3) _____ is essential. A solution that works in one country may not suit another due to differences in infrastructure, technology, or (4) _____ conditions.

One proposed strategy is to encourage households to use water responsibly. Installing (5) _____ and water-efficient appliances helps reduce usage. (6) _____ schemes can also motivate families to conserve water, linking usage directly to cost. Unfortunately, many governments don't provide enough support for (7) _____ groups.

Awareness campaigns are also vital. People often don't realize how much water is needed to produce their food, especially when it comes to (8) _____. Better education could lead to more informed consumer behavior.

On a small scale, simple practices like (9) _____ can make a significant difference. Moreover, protecting (10) _____, which naturally filter and store water, is crucial for long-term sustainability.



Please scan this QRCode for the audio file.

Answer Key:

Question no.	Answers	Question no.	Answers
1	resource management	6	metering
2	climate change	7	disadvantaged
3	Nuanced approach	8	livestock production
4	socio-cultural	9	Harvesting rainwater
5	low-flow toilets	10	natural ecosystems

5.3. Feedback from Teachers and Learners

Teacher Feedback:

- Appreciated the linguistic alignment with CEFR descriptors and realistic tone of the listening transcript.
- Noted that AI-generated summary was “pedagogically clean,” though some collocations required teacher fine-tuning.
- Recommended keeping a balance between global cohesion and local precision in paraphrasing listening summaries.

Student Feedback:

- Most students rated the task as “moderately difficult,” especially for abstract terms like *nuanced approach* or *natural ecosystems*.
- They found the topic relevant and intellectually engaging, and appreciated the clarity of the AI-generated voice.
- Suggested extending preview time before listening to 45 seconds due to the density of the summary.

5.4. Refinement Based on Feedback

Based on feedback collected during the trial run, several refinements were made:

- **Lexical clarity:** Target phrases were bolded in the teacher version of the transcript to better match the CEFR screening checklist.
- **AI prompt tuning:** Prompts were adjusted to generate more natural paraphrases and reduce overuse of academic jargon.
- **Listening script restructuring:** Minor edits were made to add rhetorical markers (e.g., “However,” “Importantly,” “For instance”) to support learners’ processing.

- **Task delivery:** Additional pre-listening support (e.g., vocabulary preview) was added for classroom use.

This case study demonstrates the feasibility and pedagogical value of recycling advanced reading texts into listening assessments using AI. The approach not only saves time but also promotes cross-skill reinforcement, contributing to a coherent and efficient test development model.

6. Discussion and Evaluation

This section synthesizes key insights from the implementation process, evaluating the strengths and limitations of integrating AI into test design for advanced learners. It also reflects on how AI-facilitated workflows affect learner autonomy and task validity, and suggests implications for future directions in language assessment development.

6.1. Strengths and Challenges of AI Use

The use of AI in designing reading and listening tests offers notable strengths. Firstly, AI significantly enhances efficiency and speed, particularly in generating question formats (e.g., gap-fill, multiple-choice), paraphrasing summaries, and producing synthetic audio for listening tasks. Tools such as ChatGPT and TTSMaker helped reduce preparation time while offering consistent linguistic quality.

Secondly, AI enables flexible content adaptation, allowing teachers to quickly adjust texts or regenerate items to suit different learner levels. The recycling of reading into listening (as shown in Section 6) illustrates how AI can support skill integration and content repurposing.

However, several challenges persist. AI-generated output often requires manual revision to refine nuance, ensure task authenticity, and eliminate distractors that are too similar or too ambiguous. Some AI tools also lack sensitivity to discourse features vital for high-level listening tasks, such as rhetorical signaling or argumentation patterns. Additionally, without teacher oversight, AI may generate items that do not align with CEFR standards or learner cognitive profiles.

6.2. Reliability and Validity Concerns

Maintaining **construct validity** is a key concern when using AI for assessment. While AI can produce linguistically complex input, it may not automatically ensure that the tasks truly reflect the cognitive demands of C1–C2 descriptors. For instance, distractors in multiple-choice tasks or gap-fill answers may not capture plausible alternatives, reducing the diagnostic value of the task.

In terms of reliability, the consistency of AI output across multiple iterations is variable. The same prompt may yield slightly different results, raising concerns about reproducibility unless prompts and versions are archived systematically. Furthermore, while automated CEFR screening tools are helpful, they cannot fully substitute for teacher judgment in identifying nuanced grammatical constructions or idiomatic expressions that influence task difficulty.

Nevertheless, by combining AI assistance with teacher validation, it is possible to achieve a high level of both reliability and validity—especially when test items undergo classroom piloting and revision.

6.3. Learner Autonomy and Test Familiarity

AI-generated tasks can contribute positively to learner autonomy, especially when students are given access to AI tools to explore, self-assess, or co-design test materials. Exposure to a variety of test types, including those generated from recycled texts, helps learners become familiar with task formats, lexical expectations, and test-taking strategies.

Moreover, the integration of AI allows teachers to offer personalized scaffolding. For example, simplified versions of listening transcripts, glossary previews, or additional AI-generated examples can be produced on demand. These features support differentiated instruction while maintaining a focus on academic-level input.

However, overexposure to AI-created tasks may create a sense of artificiality or overfamiliarity if not properly curated. Therefore, maintaining a balance between AI-assisted input and authentic, real-world materials remains essential.

6.4. Implications for Future Test Design

The findings from this study suggest several implications for future test design:

- **AI should be used as a creative assistant, not an autonomous test developer.** Human input remains critical for ensuring pedagogical alignment and psychometric soundness.
- **Recycling reading texts into listening tasks**—with AI support—presents a sustainable model for integrated skill assessment. This practice encourages cognitive transfer and allows for economical use of materials.
- **Prompt engineering and validation protocols** must be formalized. Clear prompt templates, CEFR screening tools, and revision checklists are necessary to guide AI use in a consistent and replicable manner.
- **Professional development for teachers** is key. As AI becomes more common in test design, educators need training in prompt design, output validation, and AI-literacy to make informed instructional decisions.

Ultimately, integrating AI into advanced test design opens new possibilities for scalable, learner-centered, and CEFR-aligned assessment, provided that the process remains rooted in expert judgment, reflective practice, and a commitment to quality.

CHAPTER III: CONCLUSION AND RECOMMENDATION

This final section summarizes the key outcomes of the study, translates them into actionable strategies for teachers, and outlines possible directions for future research. The integration of AI into test design for gifted English learners is shown to be both feasible and pedagogically valuable, provided that it is guided by theoretical grounding and teacher expertise.

1. Summary of Key Findings

This study has demonstrated that AI-powered tools can play a constructive role in the design of reading and listening assessments at C1–C2 levels. The case study involving the transformation of a reading passage (*“Addressing the Problem of Water Scarcity”*) into a listening summary gap-fill task highlights how AI can assist in repurposing authentic materials while maintaining CEFR alignment.

Key findings include:

- AI tools streamline the design process, especially in tasks such as paraphrasing, distractor generation, audio synthesis, and CEFR screening.
- Human intervention remains essential to refine content, ensure reliability, and preserve pedagogical coherence.
- Recycled test design, in which one source material serves multiple skill areas, enhances efficiency and supports integrated learning.
- Student feedback validates task relevance and challenge, while teacher input helps uphold quality and authenticity.

These insights confirm that AI, when used strategically and ethically, can enhance both test development workflows and learner experience in advanced-level English programs.

2. Practical Recommendations for Teachers

Based on the implementation process and evaluation outcomes, the following recommendations are offered for teachers who wish to integrate AI into their own test design practices:

1. **Curate before create:** Begin with high-quality source texts or audio, then use AI to adapt—not originate—tasks.
2. **Screen for CEFR consistency:** Use CEFR-aligned vocabulary and grammar tools (e.g., Oxford Text Checker, English Profile) to validate task level.
3. **Refine AI output collaboratively:** Use AI-generated materials as first drafts; involve teachers in reviewing, editing, and contextualizing.
4. **Re-use intelligently:** Recycle materials across modalities (e.g., reading → listening) to save time and reinforce key content.
5. **Engage students in the process:** Collect learner feedback on difficulty, clarity, and engagement, and use this to inform revisions.
6. **Build prompt libraries:** Create and refine effective AI prompts for common tasks (e.g., summary completion, inference MCQs) to ensure consistent quality.

By adopting a human-in-the-loop model, teachers can maintain control over assessment quality while benefiting from the speed and versatility of AI.

3. Suggestions for Further Research

While this study offers promising insights, several areas merit further investigation:

- Comparative studies between AI-assisted and traditionally designed assessments to evaluate performance outcomes and cognitive load on learners.
- Longitudinal research on how repeated exposure to AI-generated tasks affects learner autonomy, strategy use, and test anxiety.
- Ethical implications and bias detection in AI-generated distractors, particularly in culturally sensitive topics or high-stakes testing.
- Teacher training models focused on AI literacy and prompt engineering to support informed and critical use of AI in education.

Future research should aim to develop frameworks and protocols that support scalable, equitable, and pedagogically sound integration of AI in language assessment—especially in contexts where teachers serve gifted or high-performing student populations.

REFERENCES

1. **Alderson, J. C.** (2000). *Assessing reading*. Cambridge University Press.
2. **Buck, G.** (2001). *Assessing listening*. Cambridge University Press.
3. **Burstein, J., Tetreault, J., & Madnani, N.** (2018). The E-rater® automated essay scoring system. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of Writing Research* (2nd ed., pp. 549–563). Guilford Press.
4. **Council of Europe.** (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*.
5. **Ellis, N. C., & Winke, P.** (2020). Implications of language learning models for CEFR-aligned test development. *Language Testing*, 37(2), 251–270.
6. **Field, J.** (2008). *Listening in the language classroom*. Cambridge University Press.
7. **Godwin-Jones, R.** (2023). AI and test development: Supporting teacher-AI collaboration. *Language Learning & Technology*, 27(1), 1–15.
8. **Grabe, W., & Stoller, F. L.** (2011). *Teaching and researching reading* (2nd ed.). Routledge.
9. **Kaplan, S. N.** (2018). *Curriculum for gifted learners: Differentiation, complexity, and challenge*. Prufrock Press.
10. **Richards, J. C.** (2015). *Key issues in language teaching*. Cambridge University Press.
11. **Van der Linden, W. J.** (2019). *Computerized adaptive testing: Theory and practice*. Springer.
12. **Zhong, Z., & Meurers, D.** (2022). Leveraging large language models for test item generation: Opportunities and challenges. *Journal of Educational Technology & Society*, 25(4), 85–98.
13. **CPE Prosperity Education.** (n.d.). *Addressing the Problem of Water Scarcity* (p. 52). Prosperity Education.

14. **Scientific American.** (n.d.). *Massive Study Flips Our Story of Addiction and the Brain.* <https://www.scientificamerican.com/article/massive-study-flips-our-story-of-addiction-and-the-brain/>
15. **ChatGPT (OpenAI)** – AI language model used for drafting, refining, and generating test materials. <https://chat.openai.com/>
16. **SIDER OCR** – Optical Character Recognition tool used to extract editable text from scanned images or PDFs. <https://www.sider.ai/>
17. **Oxford Text Checker** – Tool for analyzing vocabulary by CEFR level. <https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000>
18. **TTSMaker** – Free text-to-speech tool used to generate listening input from written text. <https://ttsmaker.com/>
19. **Cathoven Level Adaptor** – A CEFR adaptation tool for adjusting reading and listening levels (experimental/limited access). <https://cathoven.com/>

APPENDICES

APPENDIX A: RECYCLED LISTENING MATERIALS

1- Original Reading Passages

ADDRESSING THE PROBLEM OF WATER SCARCITY

Despite 70% of the world's surface being covered with water, less than 3% of that is actually drinkable and, based on factors ranging from agricultural practices to climate change and daily habits, water scarcity is fast becoming a serious problem in many countries around the globe. The problem of scarcity—in other words, having insufficient water—is categorized as being physical (happening in places where supply cannot meet demand) or economic (occurring in areas that have plenty of water but don't have good management systems). Broadly speaking, the causes of water scarcity are related to the rapidly rise of the global population and the associated issues that this has brought, and the predictions in many parts of the world are that the population will continue to rise for several decades. This suggests that rather than lamenting the journey to this point, we would be better off focusing on solutions, especially given that the causes have been written about extensively and are very well understood, which is not the case for the solutions.

Solutions for water scarcity should primarily be context-specific if they are to work, and must

include experts, organisations and charities that can provide insight into the particular challenges. For example, there is no point in poor nations engaging expert engineers from overseas to help with the infrastructure for water management systems if the resulting system is not affordable or able to withstand the climatic conditions of the region. Too often collaborations on projects like this turn into vanity projects for the foreign companies in much the same way as high-end commercial projects, such as the building of a luxury hotel or impressive bridge. Workable solutions will need to understand the influences of geology, the environment and the influences of local culture to be successful and will likely be a combination of technology and a change in human behaviour. One immediate approach is to provide incentives for people to upgrade more old-fashioned machines to water-efficient ones, such as toilets and showers that use considerably less water. Another is to adjust the cost based on consumption by installing water meters in homes, something that has not been widely reported in the press. It appears that there is little incentive for governments to encourage uptake among households as any change would negatively affect voters on low incomes and families.

A better alternative to punishing such bad consumer behaviour is for governments to invest in educating their populations. For example, many are unaware of the amount of water used to produce the food we consume. The meat industry is a case in point in which vast quantities of water are required, yet the general public is largely ignorant of this. Education on how water use impacts should come from a commitment to ensuring people have the relevant scientific evidence presented to them, otherwise they will be unable to make informed decisions. There is already enough face to face floating around on the internet and it is important not to add to it.

An additional approach that is well worth thinking about is tackling the problem through multiple small lifestyles changes rather than national or international projects. After all, the situation affects millions of people, so anything that ordinary people can do without disrupting their lives too much would be a bonus. Research from behavioural science has shown that when people have to opt into a system, the likelihood of their doing so is reduced because of the increased effort involved. Rainwater collection for uses such as cleaning and washing clothes is an example of a small change. It is both low-cost and easily implementable since local councils could supply households with containers, allowing them to begin water conservation immediately.

We must also remember that better management of the environment plays a large part in maintaining the water supplies on the planet. There are certain ecosystems, such as forests, marshes and wetlands, that naturally process, collect and filter water, and preserving these natural systems is essential. Unfortunately, the practices of many common industrialised are at odds with conservation strategies for these ecosystems and so continue to be widespread. Making laws to protect these natural systems is another cost-effective way to change both attitudes and behaviour to water, and it's high time that governments stepped up and took control of the situation if we are to succeed in protecting our most precious resource.

2- Recycled Listening Test

Listening Section: Summary Completion (10 questions)

Complete the summary below using NO MORE THAN THREE WORDS from the recording.

ADDRESSING WATER SCARCITY

Water scarcity is not only a result of limited supply but also of ineffective (1) _____ and distribution systems. In some cases, the problem is not the quantity of water, but where it is and how easily it can be accessed. Influencing factors include (2) _____, population growth, and weak management.

Because each region is different, a (3) _____ is essential. A solution that works in one country may not suit another due to differences in infrastructure, technology, or (4) _____ conditions.

One proposed strategy is to encourage households to use water responsibly. Installing (5) _____ and water-efficient appliances helps reduce usage. (6) _____ schemes can also motivate families to conserve water, linking usage directly to cost. Unfortunately, many governments don't provide enough support for (7) _____ groups.

Awareness campaigns are also vital. People often don't realize how much water is needed to produce their food, especially when it comes to (8) _____. Better education could lead to more informed consumer behavior.

On a small scale, simple practices like (9) _____ can make a significant difference. Moreover, protecting (10) _____, which naturally filter and store water, is crucial for long-term sustainability.



Please scan this QRCode for the audio file.

❖ Answer key and explanation:

Question no.	Answers	Source Phrase / Paraphrase
1	resource management	<i>"it encompasses the complex dynamics of resource management and distribution"</i>
2	climate change	<i>"such as climate change, demographic pressures..."</i>
3	Nuanced approach	<i>"requires a nuanced approach tailored to specific regional contexts"</i>
4	socio-cultural	<i>"socio-cultural framework of each locality"</i>
5	low-flow toilets	<i>"such as low-flow toilets and shower systems"</i>
6	metering	<i>"implementation of metering systems that establish a direct correlation..."</i>
7	disadvantaged	<i>"particularly among economically disadvantaged populations"</i>
8	livestock production	<i>"concerning livestock production, which necessitates..."</i>
9	Harvesting rainwater	<i>"harvesting rainwater for domestic purposes..."</i>
10	natural ecosystems	<i>"efficient stewardship of natural ecosystems"</i>

3- Transcript: Addressing Water Scarcity

Did you know that although approximately 70% of the Earth's surface is covered with water, less than 3% of that is actually potable? Even then, many countries are grappling with serious water shortages. The issue of water scarcity goes beyond mere quantitative deficiency; it encompasses the complex dynamics of resource management and distribution. The challenge often lies not in the absolute availability of water resources, but rather in their geographical distribution, accessibility, and the intricate interplay of factors such as climate change, demographic pressures, and inadequate management protocols.

The implementation of effective solutions requires a nuanced approach tailored to specific regional contexts, which precludes the possibility of universally applicable strategies. The efficacy of interventions depends on the existing infrastructure, technological capabilities, and socio-cultural framework of each locality. Only through comprehensive consideration of these contextual elements can viable solutions be formulated.

One pragmatic intervention involves promoting water conservation at the household level through various mechanisms. This includes the adoption of water-efficient fixtures and appliances, such as low-flow toilets and shower systems, alongside the implementation of metering systems that establish a direct correlation between consumption and cost. However, governmental initiatives often fall short in providing adequate incentives for conservation, particularly among economically disadvantaged populations.

A critical aspect that is often overlooked is the significance of public awareness regarding water consumption. Many individuals remain unaware of the substantial water footprint associated with their dietary choices, particularly concerning livestock production, which necessitates considerable water resources. Enhanced awareness through comprehensive educational initiatives could potentially influence consumer behavior and facilitate more informed decision-making regarding water consumption.

Furthermore, the implementation of small-scale interventions can yield substantial benefits. Adopting practical solutions, such as harvesting rainwater for domestic purposes and irrigation, can significantly contribute to water conservation. These cost-effective measures can be readily integrated into daily routines, resulting in immediate positive impacts on water preservation.

Moreover, the efficient stewardship of natural ecosystems warrants attention. The intricate network of forests, wetlands, and marshlands serves as nature's own filtration and retention system, making their preservation indispensable for sustainable water management.

Ultimately, addressing water scarcity involves a combination of smart technology, education, small lifestyle changes, and the protection of nature. If we all work together, we can make a meaningful difference.

APPENDIX B: RECYCLED READING MATERIALS

1 - Original article for Reading Cloze Test design:

<https://www.scientificamerican.com/article/addiction-risk-shows-up-in-childrens-brain-scans-before-drug-use-starts/>

Massive Study Flips Our Story of Addiction and the Brain
--

Brain differences in children and teens who experiment with drugs early show up before they take their first puff or sip

For decades, Americans have been told a simple story about addiction: taking drugs damages the brain—and the earlier in life children start using substances, the more likely they are to progress through a “gateway” from milder ones such as marijuana to more dangerous drugs such as opioids. Indeed, those who start using at younger ages are much more likely to become addicted. But a recent study, part of an ongoing project to scan the brains of 10,000 kids as they move through childhood into adulthood, complicates the picture. It found that the brains of those who started experimenting with cannabis, cigarettes or alcohol before age 15 showed differences from those who did not—*before* the individuals took their first puff or sip. When paired with an independent trial of a successful prevention program tailored to at-risk kids, the findings suggest better ways to fend off substance use disorders before they start.

“This study is extremely helpful because it begins to outline the brain changes that are seen in teenagers who start to use drugs early,” says Ayana Jordan, an associate professor of psychiatry and population health at NYU Grossman School of Medicine, who was not associated with the project.

The findings are “actually telling you that there are vulnerability factors and identifying them,” says Nora Volkow, director of the National Institute on Drug Abuse (NIDA), part of the National Institutes of Health, which funded the research. Published in December 2024 in *JAMA Network Open*, the new work is part of the ongoing NIDA-led Adolescent Brain Cognitive Development initiative, the largest-ever long-term U.S. study of child brain health and growth. (Like all current NIH projects, it is threatened by the budget cuts imposed by the Trump administration, though Volkow says sustaining it is a top priority for NIDA.) In the new study, children aged nine to 11 underwent regular brain scans for three years. In separate interviews, the participants and their parents also provided information on diet and substance use. Nearly a quarter of the children had used drugs including alcohol, cannabis and nicotine before the study began.

Children who started using drugs during the study period had preexisting enlargements in many brain regions and had larger brains overall when the study began than those who did not use drugs, explains lead author Alex Miller, an assistant professor of psychiatry at Indiana University School of Medicine. These youth had many of the same brain differences as children who had begun drug use before the start of the study. In both groups, the outer portion of the brain, called the cortex, also had a larger surface area on average, with more folds and grooves.

Having a bulkier and more heavily creased brain is generally linked to higher intelligence, though these factors are far from the only ones that matter. Bigger and groovier isn’t always better: during adolescence, natural processes actually “prune back” some brain areas—so whether size differences are positive depends on the life stage being studied and on the brain regions that should be large at that time.

Other research has associated some of the brain differences found in the study with certain personality traits: curiosity, or interest in exploring the environment, and a penchant for risk-taking. Like having a large brain, curiosity and interest in novelty (which are sometimes measured together as a personality trait called “openness to experience”) are associated with intelligence. But when curiosity is coupled with a strong drive to seek intense sensations and a willingness to take risks without considering the consequences, it’s also linked to a higher likelihood of trying drugs. If these early brain differences aren’t caused by drugs, where do they come from? They could reflect certain genetic variations or childhood exposure to adverse experiences—both of which have previously been associated with addiction risk. While it’s still possible that substances could chemically interfere with brain development, contributing to the elevated risk for addiction among those who start drinking or taking other drugs early, the study suggests that there are other, preexisting factors at play.

The brain differences here were only linked to early initiation of drug *use*—not necessarily to addiction itself. “More data is needed to see if any of these brain changes are related to disease progression, severity of use or how the teens may respond to treatment,” Jordan says.

Research already suggests that early differences can be targeted to improve prevention programs. In fact, a recent trial showed that substance use disorders can be prevented in kids with personality traits that put them at higher risk. Some of the personality traits targeted in this trial have previously been associated with the kinds of brain differences found in the new brain scan study.

In the prevention trial, researchers compared Montreal-area schools in which teens received a personality-based intervention in seventh grade with those that did not. The program began by having kids take a validated personality test. Months later, with no reference to the test, teens who scored highest in the traits of impulsiveness, sensation-seeking, hopelessness or sensitivity to anxiety were invited to participate in two 90-minute workshops. These workshops taught cognitive skills aimed at maximizing the strengths and minimizing the weaknesses typically associated with their specific most strongly outlying trait.

Five years later, students at the schools that did use the program had 87 percent lower odds of developing substance use disorders. “It’s a 35 percent reduction in the annual growth of substance use disorders across time,” says Patricia Conrod, a professor of psychiatry at the University of Montreal and lead author of the prevention trial. The results were published in the *American Journal of Psychiatry* in January.

Conrod emphasizes that “risky” traits have pluses as well as minuses. For example, a tendency to seek new experiences can be critical for success in science, medicine and the arts. A willingness to take risks is useful in occupations ranging from firefighting to entrepreneurship. The trick is to help young people manage such predilections safely.

In some children she has worked with, who start drugs as early as age 13, Conrod says that “the drive to self-medicate is so strong; it’s really striking. There really is this discomfort with their inner world.” As a result, providing ways to manage these feelings without misusing drugs—and without pathologizing those with outlying traits—can be a powerful way to support healthy development.

2- Recycled Reading Cloze Test designed with AI tools:

READING COMPREHENSION

Part 1. For questions 46–55, read the following passage and fill in each of the numbered spaces with ONE suitable word.

For years, the belief that early drug use harms the brain has been deeply (46) _____ - a narrative reinforced by studies linking adolescent substance use to long-term damage. Yet a new study from the U.S. may flip that assumption. Researchers found brain differences existed even before children took their first puff or sip—suggesting certain traits may be (47) _____, not caused by drug use. These findings are (48) _____: they offer a new way of looking at addiction—not as an inevitable outcome of drug exposure, but as something shaped by biology and personality. One significant discovery was that teens who later used drugs had more folded cortices and larger brain volumes—features often (49) _____ with intelligence, risk-taking, and curiosity. Children appeared to be at their (50) _____ vulnerable when openness to experience combined with impulsivity—a mix that increased their likelihood of early experimentation. While the study itself stops short of proving causality, researchers do (51) _____ more flexible prevention strategies focused on high-risk groups. A separate school-based program showed promise. Students were tested for traits like anxiety sensitivity or sensation-seeking, and workshops were delivered months later with no direct (52) _____ to the test. The results: an 87% drop in the (53) _____ of substance use disorder over five years. Researchers stress that traits like curiosity or boldness shouldn’t be (54)

_____ ; these same features drive discovery and innovation. The challenge lies in teaching young people to channel them—without labelling the traits as problematic (55) _____ themselves.

Answer key:

Question	Best Answer	Other Acceptable Alternatives	Explanation
46	<i>ingrained</i>	<i>embedded, entrenched, accepted, widespread</i>	Collocation: “deeply ingrained belief” is most natural; others express durability or social acceptance.
47	<i>preexisting</i>	<i>inherent, innate, inborn, genetic</i>	All imply that the traits existed before drug use began.
48	<i>infectious</i>	<i>compelling, powerful, transformative, enlightening</i>	"Infectious" metaphorically fits the idea of spreading new ideas; others imply emotional or intellectual impact.
49	<i>associated</i>	<i>linked, connected, correlated, tied</i>	Standard collocations: “associated with intelligence,” etc.
50	<i>most</i>	<i>particularly, especially, extremely (less precise)</i>	“At their most vulnerable” is a fixed C1 structure. Others weaken grammatical precision but retain meaning.
51	<i>recommend</i>	<i>propose, advocate, support, suggest</i>	All are valid verbs of recommendation in academic contexts.
52	<i>reference</i>	<i>mention, link, relation, connection</i>	“Reference to the test” is the most idiomatic; others are acceptable with a slight shift in nuance.
53	<i>odds</i>	<i>likelihood, rate, risk, chance(s), incidence</i>	“Drop in the odds” is idiomatic; others are statistically acceptable and contextually logical.
54	<i>stigmatized</i>	<i>demonized, criticized, pathologized, shamed</i>	All reflect negative social judgment of traits.
55	<i>in</i>	<i>/./</i>	“In themselves” is a C1 academic phrase, focuses on essential nature.

APPENDIX C: QUALITY CONTROL

a. CEFR Lexico-Grammar Checklist

Section 1: Test Overview

Item	Description
Test Name	
Language Skill	☐ Reading ☐ Listening
CEFR Level	☐ C1 ☐ C2
Author	
Date Created / Revised	
Number of Items	
Total Time	
Task Types	e.g., MCQs, T/F/NG, matching, etc.

Text/Audio Title(s)	
Text/Audio Source	☐ Teacher-generated ☐ Adapted

Section 2: Evaluation Rubric

1. CEFR Alignment and Language Appropriateness

Criterion	Yes	No	Notes
Lexical range appropriate for C1–C2 (based on EVP, English Profile)	☐	☐	
Use of advanced grammar structures (modality, inversion, complex clauses)	☐	☐	
Text includes idioms, collocations, or abstract vocabulary	☐	☐	

2. Cognitive Demand and Task Authenticity

Criterion	Yes	No	Notes
Requires inference beyond surface-level reading/listening	☐	☐	
Includes tasks similar to real-life academic or professional contexts	☐	☐	
Topics are appropriate for mature, educated learners	☐	☐	

3. Quality of Text/Audio Input

Criterion	Yes	No	Notes
Text/audio is coherent, well-structured, and stylistically consistent	☐	☐	
Audio includes natural features: hesitation, intonation, variation, etc.	☐	☐	
No grammatical or stylistic awkwardness in artificial materials	☐	☐	

4. Variety and Depth of Question Types

Criterion	Yes	No	Notes
Includes global and local comprehension questions	☐	☐	
Questions test inference, reference, tone, vocabulary meaning in context	☐	☐	
Distractors are plausible and not guessable	☐	☐	

5. Reliability and Scoring

Criterion	Yes	No	Notes
Clear answer key or transcript provided	☐	☐	
Each item targets a specific sub-skill	☐	☐	
Answer choices are balanced and free from trick questions	☐	☐	

6. Accessibility and Bias

Criterion	Yes	No	Notes
No culturally biased or exclusive content	☐	☐	

Instructions are clear and unambiguous	☐	☐	
Layout is learner-friendly (font size, spacing, audio clarity)	☐	☐	

Section 3: Trialing & Feedback

Question	Response
Has this test been trialed with learners?	☐ Yes ☐ No
If yes, what changes were made post-trialing?	
Any student/peer feedback collected?	☐ Yes ☐ No

b. Sample Student Feedback Sheet

Title: Student Feedback on AI-Generated Reading and Listening Test (Trial Version)

Target Group: Gifted Grade 11 English Majors – Pilot Trial Participants

Instructions: *Please answer the questions below honestly.*

Your feedback will help improve future test design using AI-generated materials.

Section 1: Listening Task – Technical Aspects

1. Was the voice in the listening task clear and easy to understand?
 - ☐ Very clear and natural
 - ☐ Clear but slightly unnatural
 - ☐ Difficult to follow
2. How would you rate the speed of the speaker?
 - ☐ Too fast
 - ☐ Just right
 - ☐ Too slow
3. Did the audio include enough pauses before important information?
 - ☐ Yes, the pacing was helpful
 - ☐ Sometimes, but could be improved
 - ☐ No, the key points were too close together

Section 2: Language Difficulty

4. How difficult was the vocabulary in the reading and listening texts?
 - ☐ Very difficult – I didn't know many words
 - ☐ Moderately difficult – I needed to pre-learn a few words
 - ☐ Easy – I understood most of it
5. How would you describe the grammar structures in the listening task?
 - ☐ Too complex – some sentences were hard to follow
 - ☐ Acceptable for my level
 - ☐ Too simple
6. Do you think a short glossary (3–5 words) before the listening would help?
 - ☐ Yes
 - ☐ No

Section 3: Content and Design

7. Did you find the topics of the tasks interesting?
 - ☐ Very interesting and relevant
 - ☐ Neutral
 - ☐ Boring or outdated
8. How did you feel about the summary completion task in the listening test?

- ☐ Clear and well-designed
- ☐ Too difficult
- ☐ Too easy or guessable
- 9. Were there any answers you feel were incorrect or misleading in the answer key?
 - ☐ Yes → If yes, which one(s)? _____
 - ☐ No
- 10. Was the time allowed for the listening section enough?
 - ☐ Yes, I finished comfortably
 - ☐ Almost enough – I needed 1–2 more minutes
 - ☐ No – I needed more than 3 minutes extra

Section 4: Suggestions

- 11. Which of the following would you like to see improved in future versions? (*Tick all that apply*)
 - ☐ Use more natural or human voices
 - ☐ Include a short glossary before listening
 - ☐ Balance grammar and vocabulary better
 - ☐ Make gap-fill answers harder to guess
 - ☐ Add more listening time
 - ☐ Others (please specify): _____
- 12. Overall, how would you rate the AI-generated tasks?
 - ☐ Excellent – useful and realistic
 - ☐ Good – with some areas to improve
 - ☐ Fair – needs several changes
 - ☐ Poor – not suitable for testing

Thank you for your time and input! Your answers will help us refine the tests for future use.

Student Feedback Summary

Feedback Category	Summary of Responses
Clarity of Audio	7/12 reported the voice was clear but “less natural” than human speakers
Speed of Listening	3 found it too fast; 2 suggested clearer pauses before answer points
Lexical Difficulty	9/12 felt the vocabulary was “dense” and required pre-teaching of 3–5 words
Grammar Complexity	Acceptable but some sentences were overly long or difficult to process orally
Engagement with Topic	All agreed the topics were “interesting and modern”
Task Design (Summary Completion)	4 noted that some gaps were “guessable” based on collocations rather than actual audio
Answer Key Accuracy	2 errors in suggested answer key were found by students during discussion
Time Appropriateness	3 students needed extra 2–3 minutes to finish listening section
Suggestions from Students	Use varied voices; include pre-listening glossary; balance

	grammar vs. vocabulary demand
--	-------------------------------