



Constructing CEFR-Aligned English Reading Assessments with AI: A Practical Guide for Thai English Teachers in Higher Education

Sun-Young Shin^a, Suchada Sanonguthai^{b,*}

^a shin36@iu.edu, Department of Linguistics, Second Language Studies, Indiana University Bloomington, USA

^b suchada.s@litu.tu.ac.th, Language Institute, Thammasat University, Thailand

* Corresponding author: suchada.s@litu.tu.ac.th

APA Citation: Shin, S.-Y. & Sanonguthai, S. (2026). Constructing CEFR-aligned English reading assessments with AI: a practical guide for Thai English teachers in higher education. <i>LEARN Journal: Language Education and Acquisition Research Network</i> , 19(2), 71-155. https://doi.org/10.70730/FDFL1823	
Received 02/12/2025	ABSTRACT This paper provides practical guidance for developing English reading assessments aligned with the Common European Framework of Reference for Languages (CEFR) with the support of generative AI. Although the CEFR is widely used in curriculum design, instruction, and assessment, many instructors face challenges when translating reading descriptors into assessment tasks that appropriately reflect specific proficiency levels, particularly transitional levels such as B1+ and B2+. AI tools offer new opportunities for generating reading materials and assessment tasks, but recent studies show that CEFR level labels alone do not ensure appropriate alignment. Drawing on language assessment theory and recent research on AI-generated CEFR-aligned texts, this paper proposes a specification-driven approach in which reading test specifications are derived from CEFR-aligned instructional materials and target CEFR descriptors to guide AI-assisted assessment development through a three-stage prompting process consisting of reading test specification development,
Received in revised form 15/05/2026	
Accepted 01/06/2026	

	reading assessment generation, and optional review and revision. For illustrative purposes, a baseline prompting approach based on CEFR-aligned textbook tasks and target CEFR descriptors is also presented to represent a realistic classroom assessment development process. Illustrative applications at the B1, B1+, and B2 levels show how reading test specifications can be incorporated into AI prompts to guide the development of CEFR-aligned reading assessments. The paper concludes with practical guidance for adapting AI-assisted assessment development to classroom contexts while maintaining instructor judgment and alignment with the intended reading construct. Keywords: CEFR, reading assessment, AI, Thai higher education, assessment literacy, educational technology
--	---

Introduction

The Common European Framework of Reference (CEFR; Council of Europe, 2001, 2020) has become a widely used reference framework for English language teaching and assessment across Southeast Asia. In Thailand specifically, university exit examinations, graduation benchmarks, and institutional English programs increasingly invoke CEFR levels as standards of expected achievement (Charttrakul & Damnet, 2021; Wudthayagorn, 2022). Despite this widespread adoption, the CEFR tends to function as a framework of aspirational labels rather than actionable criteria in Thai higher education, particularly where formal training in language testing is not readily available to English language teachers. This gap leads to a persistent misalignment between stated CEFR level targets and the actual cognitive and linguistic demands of assessments in practice.

Research on English language assessment in Thai higher education identifies several recurring challenges. Although many Thai university instructors report awareness of the CEFR, studies indicate that practical understanding of how to apply its descriptors to teaching and assessment remains limited, suggesting a need for additional training and implementation support (Charttrakul & Damnet, 2021; Phoolaikao & Sukying, 2021). A related issue concerns students: research has revealed discrepancies between Thai university students' self-assessed CEFR levels and their externally measured proficiency (Piamsai, 2023). Such discrepancies likely reflect the broader difficulties instructors face in implementing CEFR-aligned assessment practices and in interpreting descriptors consistently across educational contexts.

Two features of current practice intensify these challenges. First, the CEFR encompasses not only its six primary levels (A1 through C2) but also a set of "plus levels" (e.g., A2+, B1+, and B2+) that provide finer-grained benchmarks at transitional stages of proficiency. These levels are further elaborated in the Companion Volume (Council of Europe, 2020) but are rarely addressed in practitioner-oriented guides. Second, the emergence of large language models (LLMs) as accessible AI content-generation tools (e.g., ChatGPT, Claude, and Gemini) presents both an opportunity and a methodological challenge: the extent to which AI-generated materials reliably respect CEFR level constraints remains an open question.

This paper responds to both challenges by proposing a specification-driven approach for AI-assisted reading assessment development. The approach uses CEFR-aligned instructional materials as starting points for deriving reading test specifications. These specifications are then incorporated into a three-stage AI prompting process. Rather than presenting an empirical evaluation or comparing the effectiveness of different AI prompting approaches, this paper serves as a practice-oriented methodological guide that illustrates the proposed specification-driven approach through three illustrative applications at the B1, B1+, and B2 levels. The paper includes examples at the B1, B1+, and B2 levels based on reading tasks derived from commercially published CEFR-aligned textbooks available in Thailand. The paper also demonstrates how instructors may use tools such as Text Inspector (Bax, n.d.) as part of the review process when evaluating AI-generated materials. Finally, practical guidance is provided to help Thai university instructors incorporate AI into reading assessment development while maintaining alignment with instructional objectives and applying professional judgment when reviewing AI-generated materials.

Theoretical Background

The CEFR and Reading Assessment

The Companion Volume extends and refines many CEFR descriptors, particularly for reading, and formalizes the plus levels as legitimate intermediate benchmarks. These benchmarks are especially relevant to the Thai higher education context. Most students enter university at approximately A2 (Waluyo, 2019), while national guidelines set B1 as the expected proficiency for Grade 12 completers and many universities target B2 as the exit level for non-English majors (Wudthayagorn, 2022). This entry-to-exit span means that a substantial proportion of learners pass through the intermediate CEFR bands during their university studies, making the plus

levels (A2+, B1+, and B2+) particularly consequential for both instructional planning and reading assessment design.

Reading ability is also a central academic skill for students in Thai higher education more broadly. University students are expected to read course materials, online information, research summaries, professional documents, institutional announcements, and, increasingly, AI-mediated texts in English. For many students, reading also underpins other academic activities, including notetaking, presentation preparation, and discipline-specific writing. In this context, reading assessment is not simply a classroom procedure for assigning grades; it is a means of clarifying learning goals, monitoring progress, supporting instruction, and enabling institutions to make defensible claims about students' English proficiency.

The CEFR offers a useful reference point for this work because it describes language proficiency through communicative activities and can-do descriptors rather than through test scores alone. The Companion Volume emphasizes that the framework is intended to support coherence among learning, teaching, and assessment, and that its descriptors can help teachers connect classroom goals with real-world language use (Council of Europe, 2020). For reading, this orientation is especially important. A CEFR-aligned reading test should not merely examine whether students know isolated vocabulary or grammar; it should assess whether students can use written input for meaningful purposes, such as identifying relevant information, understanding main ideas, following arguments, interpreting instructions, and engaging with texts related to their academic or professional needs.

This relevance is particularly acute in Thailand, where higher education institutions have increasingly adopted CEFR levels as benchmarks for English learning outcomes. Siripol et al. (2025) note that Thai higher education policy has encouraged institutions to align English learning and assessment practices with CEFR descriptors, including expectations that undergraduate and graduate students reach relatively high levels of English proficiency. Such policy developments create an urgent practical need: teachers must understand how to translate broad CEFR levels into classroom-level reading tasks and test specifications that are feasible, transparent, and locally appropriate.

This need is further intensified by the rapid spread of generative AI. AI tools can help teachers draft reading passages, produce multiple item types, adjust text length, and generate explanations. Such affordances are particularly attractive in contexts where teachers have limited time to prepare materials across different CEFR levels. However, recent studies caution against treating AI as an automatic alignment solution. Uchida (2025), for example, found that although ChatGPT-4o generated texts that differed on the surface by CEFR level, they did not consistently match CEFR-based

textual features: lower-level texts tended to be overly simple, while higher-level texts were often too difficult. The study also identified topic bias, with A-level texts frequently focusing on daily life and more advanced texts gravitating toward sustainability, environmental issues, or artificial intelligence. Similarly, Siripol et al. (2025) found substantial inconsistency among automated CEFR analyzers when classifying B1 and B2 descriptive texts, suggesting that such tools may produce divergent judgments for the same text depending on their algorithms, metrics, and assumptions.

Nguyen et al. (2026) provide further evidence on this point from a vocabulary load perspective. Analyzing over four million tokens of ChatGPT-generated texts across six CEFR levels using Nation's (2012) BNC/COCA frequency lists, they found that while the overall lexical demand of AI-generated texts increased with CEFR level, important mismatches remained. Texts targeting A1 and A2 required the same vocabulary size for optimal comprehension as B1 texts (3,000-word families), meaning that the lower-level texts were lexically more demanding than they should have been for those learner groups. At the other extreme, C2-level texts required fewer word families than Nation's estimates for C2-proficient readers would predict. The study also found that B1-level texts had the highest coverage of high-frequency words and the lowest coverage of mid-frequency words across all levels. This pattern runs counter to what a moderate-difficulty B1 text would be expected to show. Nguyen et al. attribute many of these discrepancies to the prompts used: when ChatGPT was not constrained to generate topic-diverse texts, it defaulted to narrow topic ranges at each level (e.g., daily life for lower levels, abstract or technical topics for higher levels), which skewed the resulting lexical profiles. Their study concludes that while ChatGPT can generate texts that are broadly differentiated by level, human verification of lexical features against CEFR-based benchmarks remains necessary before using AI-generated materials in instructional or assessment contexts.

Taken together, these findings point to a gap in the existing literature that the present paper directly addresses. All three studies cited above examined AI-generated texts produced from relatively minimal prompts that specified a CEFR level but did not include structured test specifications. None of the studies investigated whether providing ChatGPT with detailed assessment design information would produce texts that more closely match CEFR descriptors and are more suitable for classroom assessment. This is a meaningful gap, because in principled assessment development, the test specification is the instrument that translates a broad proficiency level into concrete design constraints. It is therefore reasonable to explore whether AI prompts that incorporate these constraints would produce assessment-ready

materials that are more closely aligned with CEFR descriptors and more suitable for classroom assessment.

The purpose of this guide is not to promote AI as a replacement for teacher judgment. Rather, it argues that AI can be most useful when used within a principled assessment design process. For Thai higher education teachers, this means beginning with a clear reading construct and relevant CEFR descriptor, using AI to support passage and item generation, and reviewing the resulting materials in light of local curricular needs and assessment goals. This observation raises an important practical question: if CEFR descriptors alone are insufficient for guiding assessment development, what additional information is needed to translate broad proficiency descriptions into concrete reading tasks and assessment materials? In language assessment, this function is typically served by test specifications. The next section therefore examines how reading test specifications can be used to operationalize CEFR descriptors and support AI-assisted assessment development.

CEFR Reading Descriptors as a Basis for Assessment Design

The CEFR reading descriptors provide a practical starting point for assessment design because they describe what learners at different levels can typically do with written texts. In the Companion Volume, reading comprehension is classified under reception and defined broadly to include the processing of written or signed texts (Council of Europe, 2020). This broad view is useful for classroom assessment because it foregrounds that reading ability involves more than decoding words; it involves drawing on textual, contextual, and background knowledge to construct meaning for a specific purpose.

The level descriptors also guide principled decisions about text difficulty. At A1, learners can understand very short, simple texts, often one phrase at a time, by recognizing familiar names, words, and basic phrases. At A2, learners can understand short, simple texts on familiar, concrete matters, especially when vocabulary is high frequency and the text is predictable. The A2+ level can be understood as a stronger form of A2 performance. This suggests that A2+ texts may still need to be short, concrete, and familiar, but they can require more integration of information across sentences, more recognition of simple text organization, and somewhat greater tolerance of detail than A2 tasks.

The distinctions become particularly important at the intermediate levels, where learners move beyond understanding straightforward information to engaging with increasingly complex texts and reading demands. At B1, learners can read straightforward factual texts on familiar or

personally relevant subjects with satisfactory comprehension. The B1+ level represents a “strong Threshold” performance. In other words, B1+ tasks may move beyond locating main ideas and details in straightforward texts to requiring learners to summarize the main point of a short article, identify a writer’s opinion, explain why something is presented as a problem, or answer follow-up questions about specific details. B1+ is especially important for transitional learners because it marks the point at which reading tasks can begin to include more extended informational texts, clearer argumentative structure, and moderate inferencing while still remaining anchored in familiar topics and transparent organization.

At B2, learners can read with a large degree of independence, adapting their reading speed and style to different purposes and using reference sources selectively. In contrast, B2+ tasks require more sustained engagement with argumentation, comparison of viewpoints, recognition of discourse relationships, and evaluation of how ideas are linked across a text. Whereas B2 readers may understand the main ideas and supporting details of complex texts, B2+ readers can be expected to handle denser argumentative organization, identify relationships between claims and evidence, and follow more developed lines of reasoning, provided the text remains within a reasonable range of academic or professional familiarity. At C1, learners can understand lengthy and complex texts, including academic and professional publications, especially when rereading is possible. These distinctions illustrate that level is not determined solely by passage length; it also depends on text type, topic familiarity, lexical range, syntactic complexity, information density, discourse organization, and the degree of inference required.

While the descriptors provide a valuable design foundation, they should not be applied mechanically. Descriptors are illustrative, not prescriptive test specifications, as they indicate what learners can do, but they do not automatically determine the passage, item format, scoring procedure, or acceptable response. For example, the B1 descriptor for reading straightforward factual texts can be realized through many possible tasks: identifying the main idea of a short article, matching headings to paragraphs, answering selected-response questions about details, or completing a table with information from a text. Each task may be appropriate if it elicits the intended reading behavior, but each also introduces different demands. Matching headings may require discourse-level understanding, while scanning for details may require only locating specific words or phrases. A CEFR-based assessment must therefore connect the descriptor to both the intended cognitive process and the chosen item format.

This point becomes especially important when AI is involved. A generative AI tool can produce a passage labeled “B1” or “B2,” but the label itself is not evidence of alignment. Uchida (2025) demonstrates that

AI-generated texts may display surface-level differences by level while still diverging from CEFR-based benchmarks in vocabulary, readability, and topic distribution. Siripol et al. (2025) further show that automated CEFR analyzers do not always agree with one another, even when classifying the same texts. Together, these findings suggest that teachers should treat AI outputs and automated CEFR classifications as drafts or advisory evidence rather than final decisions. Human review remains necessary to judge whether the text, task, and item demands correspond to the selected descriptor and to learners' local context.

From Descriptors to Reading Test Specifications

A reading test specification is a short planning document that describes what the test is intended to measure and how it will do so. It functions as a bridge between a general CEFR descriptor and the actual test materials. Without a specification, teachers may rely too heavily on intuition when selecting passages or writing items, producing tests that are convenient but only loosely connected to the intended CEFR level. A specification helps ensure that the passage, items, scoring, and score interpretation are all aligned with the reading construct.

The need for test specifications is also supported by broader concerns about proficiency scales. Shin (2026) argues that although proficiency descriptors promote transparency and coherence, they do not always specify how contextual factors, task features, and content selection affect actual performance. Test specifications therefore play a necessary role in translating broad proficiency descriptors into concrete assessment decisions about text type, topic, item format, cognitive demand, and scoring.

The value of test specifications extends beyond classroom assessment design. As the literature on AI-generated reading materials has begun to show, the quality and CEFR alignment of AI outputs depends substantially on the information and constraints provided during generation. Recent research on automatic item generation has similarly found that prompt-based items may show inconsistent quality and weak construct alignment, highlighting the importance of explicitly defining targeted language subskills and principled item-design criteria (Aryadoust & Wong, 2026). Uchida (2025), Siripol et al. (2025), and Nguyen et al. (2026) also found that ChatGPT-generated texts showed meaningful deviations from CEFR expectations using prompts that specified little more than a target CEFR level. However, none incorporated the kind of structured assessment design information that a test specification would supply. This raises an important practical question: if AI prompts are enriched with the full content of a test specification including the target construct, text type, topic domain,

vocabulary range, syntactic expectations, and intended cognitive processes, would the resulting texts be more closely aligned with CEFR descriptors and more suitable for use in reading assessment?

The elements of a test specification are described in the following subsections. For each element, the discussion notes both its role in classroom assessment design and its relevance to AI-assisted text generation. The first element of a specification is the target CEFR level, selected on the basis of course goals, program expectations, or student profile. In Thai higher education, a general English course may focus on A2+ to B1 outcomes, while an English for academic purposes course may target B1+ to B2 reading. A graduate-level academic reading course may target B2+ or C1, depending on the discipline and institutional expectations. Teachers should avoid assuming that all students in a course operate at the same level; when proficiency varies widely, the test may include tasks at more than one adjacent level or use a common passage with items that vary in cognitive demand. For AI-assisted text generation, stating the target level explicitly in the prompt is a necessary but insufficient starting point: as Nguyen et al. (2026) show, a level label alone does not prevent ChatGPT from producing texts with vocabulary loads that are mismatched to that level's learner profile.

The second element is the reading construct, which is a statement of what aspect of reading ability is being assessed. A clear construct might be "reading for specific information in short institutional texts," "understanding the main idea and supporting details in a factual article," or "identifying claims, reasons, and writer stance in a short academic text." Construct clarity prevents the test from becoming a general language quiz. If the construct is reading for argument, items should focus on relationships among ideas rather than vocabulary definitions. If the construct is reading instructions, the task should require students to understand procedures, conditions, or warnings. Including the construct in an AI prompt also constrains the output in ways that a level label cannot: a prompt specifying "generate a B1 reading passage in which the main comprehension demand is following a sequence of instructions" will typically yield a more focused and level-appropriate text than a prompt specifying only "generate a B1 reading passage."

The third element is text type. CEFR descriptors reference a range of texts, including notices, advertisements, letters, brochures, websites, reports, articles, instructions, and academic or professional publications. Text type should be chosen to reflect students' likely reading needs. Thai university students may encounter course announcements, scholarship information, research abstracts, workplace emails, public health notices, tourism materials, short opinion articles, textbook passages, or discipline-related webpages. For AI-supported test development, teachers should specify the text type clearly in the prompt rather than asking AI to generate a generic passage. This is

particularly important given the topic bias identified by Nguyen et al. (2026) and Uchida (2025): without explicit guidance, ChatGPT tends to default to certain topic and text-type combinations at each CEFR level, which can skew lexical profiles and reduce the representativeness of the resulting materials.

The fourth element is topic domain. Topics should be familiar enough for the target level without being so familiar that students can answer correctly without reading the text. At lower levels, topics might involve daily routines, campus life, travel, food, health, or simple workplace situations. At B1 and B2, appropriate topics include university study, technology use, environmental issues, intercultural communication, public transportation, health information, and workplace practices. At C1, topics may be more abstract or specialized such as academic integrity, sustainability policy, public health communication, or the social implications of AI. Topic selection should also consider local relevance. In Thai higher education, locally meaningful topics can improve engagement and reduce cultural mismatch, provided that the task still requires comprehension of the text rather than background knowledge alone. Nguyen et al.'s (2026) finding that prompt-specified topic diversity significantly improved the lexical representativeness of ChatGPT-generated corpora underscores the importance of making topic domain an explicit element of any AI generation prompt.

The fifth and sixth elements are text length and language expectations. Text length should be appropriate for the level, purpose, and testing time. A2 tasks may use short texts of approximately 80 to 150 words, or several very short texts. B1 tasks may use texts of approximately 180 to 350 words, B2 tasks approximately 350 to 600 words, and C1 tasks may use longer and denser texts if the testing purpose requires it. These ranges should be treated flexibly, since length alone does not determine difficulty. Vocabulary and grammar expectations should also be specified. A B1 text may include mostly high-frequency vocabulary, familiar academic words, clear cohesive devices, and straightforward sentence structures. A B2 text may feature more varied vocabulary, less frequent academic words, embedded clauses, and more complex discourse markers. Stating these expectations explicitly in the prompt can provide the AI model with concrete linguistic constraints that go well beyond a level label. For example, instructors may ask AI to prioritize high-frequency vocabulary and limit low-frequency words, while independently reviewing whether the resulting lexical demands are appropriate for the intended learners and task.

The seventh and eighth elements are item format and number of items. Common formats include multiple-choice questions, matching, short answer, information transfer, true/false with justification, gap completion, and ordering. Each format has strengths and limitations: multiple-choice items are easy to score but difficult to write well; short-answer items reduce

guessing but require clear scoring rules; matching headings to paragraphs assesses discourse comprehension but can become confusing if options overlap; information-transfer tasks, such as completing a table, are well suited to reading for orientation or extracting specific information. The number of items should be sufficient to sample the construct but not so large that the test becomes repetitive. A short classroom task may include three to five items, while a formal quiz may include ten to fifteen items across two or three texts.

The ninth element is the intended cognitive process. Teachers should identify whether students are expected to locate information, identify main ideas, infer meaning, understand text organization, interpret writer stance, compare information, or evaluate an argument. This element is essential for CEFR alignment because the same passage can support tasks at different levels: a B2-level text with simple detail questions may function as a lower-level task, while a B1-level text with complex inference questions may become more demanding than intended. The cognitive process must therefore be aligned with both the descriptor and the expected level.

The final element is the scoring method. Selected-response items may be scored using an answer key, while constructed-response items require scoring criteria. For short-answer items, acceptable responses should be specified in advance, including whether spelling or grammar errors affect credit. For tasks requiring justification, partial credit may be appropriate when students identify the correct answer but provide incomplete evidence. In classroom use, scoring can also incorporate feedback categories such as locating information, understanding main ideas, interpreting vocabulary from context, and recognizing writer stance as this diagnostic feedback is often more useful than a single total score. Table 1 presents a simple test specification template that teachers can adapt before using AI to generate passages or items.

Table 1

Sample CEFR-Aligned Reading Test Specification Template

Specification element	Guiding question or decision
Target CEFR level	Which level or adjacent level range is the test intended to represent?
Reading construct	What reading ability is being assessed, such as scanning, main idea comprehension, argument comprehension, or reading instructions?

Specification element	Guiding question or decision
Text type	What genre or format will students read, such as an announcement, webpage, article, email, report, or instructions?
Topic domain	Is the topic personal, public, educational, occupational, academic, or discipline-specific?
Text length	How long should the text be for the level, purpose, and time limit?
Vocabulary and grammar expectations	What range of vocabulary, sentence structure, cohesion, and discourse organization is appropriate?
Item format	Which item type best elicits the intended reading behavior?
Number of items	How many items are needed to sample the construct without overburdening students?
Intended cognitive process	Will students locate, identify, infer, compare, interpret, or evaluate information?
Scoring method	Will responses be scored with an answer key, rubric, partial credit rule, or diagnostic categories?

Example Specification

For a B1 university general English class, a teacher might define the construct as reading a campus announcement to locate relevant information and understand the main purpose. The text type could be a 220-word university webpage announcement about a volunteer program. Items could include two detail questions, one main purpose question, and one information-transfer table. Scoring could use an answer key with one point per item. For a B2 academic reading class, the construct might shift to identifying claims and supporting reasons in a 450-word opinion article about AI use in education, with items asking students to identify the author's position, distinguish evidence from opinion, and infer the function of a paragraph.

The discussion above suggests that test specifications can play an important role in improving the quality and CEFR alignment of AI-generated reading materials. However, the practical question remains how these specifications can be incorporated into AI prompting in a systematic and

replicable manner. The next section addresses this question by presenting a specification-driven approach for AI-assisted reading assessment development. Using CEFR-aligned textbook chapters as starting points, the approach demonstrates how reading test specifications can be derived from existing materials and subsequently used to guide the generation of reading passages and assessment tasks at the B1, B1+, and B2 levels.

A Specification-driven approach for AI-Assisted Reading Assessment Development

Building on the principles discussed in the previous section, the proposed specification-driven approach operationalizes CEFR descriptors through explicit reading test specifications before AI is used to generate assessment materials. Rather than prompting an AI system to generate a reading assessment using only a target CEFR level, the approach begins by analyzing a CEFR-aligned textbook chapter together with the corresponding target CEFR descriptor to develop a reading test specification. The completed specification then serves as the blueprint for subsequent assessment development by defining the reading construct, text type, topic domain, text length, vocabulary and grammar expectations, item format, intended cognitive process, and scoring method before AI generates a reading passage or assessment tasks.

The approach consists of three stages (Figure 1). Stage 1 focuses on developing a reading test specification from a CEFR-aligned textbook chapter and a target CEFR descriptor. Stage 2 uses the completed specification to generate a new reading assessment consisting of a reading passage, assessment items, and an answer key. Stage 3 provides an optional review and revision stage that allows teachers to evaluate and refine the generated assessment while maintaining consistency with the intended reading construct and target CEFR level. Together, these three stages place assessment design at the center of AI-assisted assessment development by ensuring that key assessment decisions are established before assessment materials are generated. For the illustrative applications presented in this paper, all prompts were submitted to the GPT-5.5 model via ChatGPT (OpenAI, n.d.). The resulting outputs were used solely to illustrate the proposed specification-driven approach.

Figure 1

Overview of The Three-Stage Specification-Driven Approach for AI-Assisted Reading Assessment Development

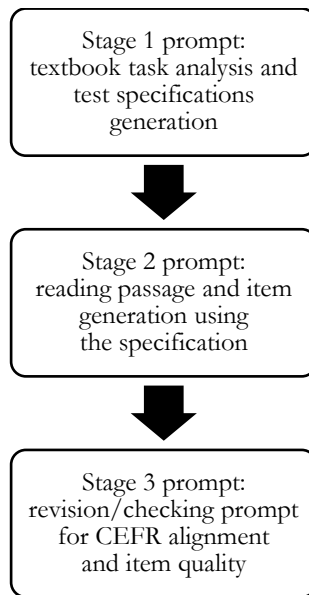


Figure 1 summarizes the overall workflow of the proposed approach. The complete prompts supporting Stages 1, 2, and 3 are presented in Appendices A, B, and C, respectively. To illustrate the contribution of the proposed approach, a baseline prompting approach is also included, with the complete baseline prompt presented in Appendix D. Illustrative applications of both the specification-driven and baseline approaches at the B1, B1+, and B2 levels are presented in Appendices E, G, and I and Appendices F, H, and J, respectively. Appendix K illustrates how one of the optional review and revision prompts can be used to refine a generated assessment while maintaining consistency with the intended reading construct and target CEFR level. The following subsections describe each stage of the proposed approach before illustrating its application across the three CEFR levels.

Stage 1: Reading Test Specification Development

The first stage establishes the reading test specification that guides the subsequent stages of assessment development. As discussed in Section 3, reading test specifications translate broad CEFR descriptors into explicit assessment characteristics by defining the reading construct, text type, topic

domain, text length, vocabulary and grammar expectations, item format, intended cognitive process, and scoring method. Rather than asking AI to generate a reading assessment directly from a target CEFR level, the proposed approach first develops these assessment characteristics so that assessment decisions are made before any reading materials are generated.

To support this process, a CEFR-aligned textbook chapter and a target CEFR descriptor are analyzed using Prompt A and the specification framework presented in Table 1. Instead of reproducing the original textbook task, the prompt examines the chapter as a whole and draws on evidence from its reading texts, reading activities, and learning objectives to develop a reading test specification. The resulting specification is intended to support the development of a new parallel reading assessment while maintaining alignment with both the instructional objectives represented in the textbook chapter and the selected CEFR descriptor.

In addition to the completed reading test specification, Prompt A generates a Summary of Key Assessment Characteristics that synthesizes the principal assessment decisions identified during the specification-development process. This summary provides a concise reference to the target CEFR level, reading construct, text type, topic domain, text length, vocabulary and grammar expectations, item format, number of items, intended cognitive process, and scoring method. Rather than introducing new information, it summarizes the completed specification and serves as the primary input for the assessment generation stage.

The complete Prompt A is presented in Appendix A, while illustrative applications for the B1, B1+, and B2 textbook chapters are provided in Appendices E, G, and I, respectively. Together, these applications demonstrate how the same specification-development procedure can be applied to different CEFR levels using commercially published CEFR-aligned instructional materials as the starting point. Once the reading test specification has been established, it provides the basis for generating a new CEFR-aligned reading assessment in the second stage of the proposed approach.

Stage 2: Reading Assessment Generation

Once the reading test specification has been established, the second stage focuses on generating a new CEFR-aligned reading assessment. Whereas Stage 1 identifies the assessment characteristics that define the intended construct, Stage 2 translates those characteristics into assessment materials. Rather than relying on a target CEFR descriptor alone, assessment generation is guided by the completed reading test specification, ensuring that

the resulting assessment is developed from explicit assessment decisions rather than a proficiency label in isolation.

Assessment generation is carried out using Prompt B, which incorporates both the completed reading test specification and the Summary of Key Assessment Characteristics produced during Stage 1. Together, these provide ChatGPT with the information needed to generate an original reading passage, assessment items, an answer key, and brief explanations for the correct responses. Because the prompt is informed by the completed specification, the generated assessment is intended to reflect the reading construct, text type, topic domain, language expectations, item format, and intended cognitive processes established during the specification-development stage.

An important feature of this stage is that the generated assessment functions as a parallel assessment rather than a reproduction of the original textbook task. Although the textbook chapter serves as the starting point for specification development, the assessment generated by AI is newly created and is designed to reflect the same assessment characteristics without copying the original instructional materials. In this way, the completed specification provides continuity between the source materials and the new assessment while allowing instructors to develop additional classroom assessments aligned with the same learning outcomes.

As with any assessment development process, instructors may decide to revise the reading passage, assessment items, or task design to better suit their instructional objectives, learners, or local teaching context. The proposed approach therefore includes an optional review and revision stage to support this final step in the assessment development process. The complete Prompt B is presented in Appendix B, and illustrative applications for the B1, B1+, and B2 textbook chapters are provided in Appendices E, G, and I, respectively.

Stage 3: Assessment Review and Revision

Although the assessment generated during Stage 2 is designed to reflect the completed reading test specification, instructors may still wish to review and refine the materials before classroom use. The third stage therefore provides an optional review and revision process that supports the evaluation of AI-generated assessments while maintaining consistency with the intended reading construct and target CEFR descriptor. Depending on the instructional context, instructors may wish to adapt the assessment to accommodate practical considerations while preserving the assessment characteristics that define the intended construct.

To support this process, Prompt C consists of three optional review and revision prompts that may be used independently or in combination. Prompt C1 provides a structured review of the generated assessment by evaluating its alignment with the completed reading test specification and identifying aspects that may warrant revision. Rather than revising the assessment automatically, it encourages instructors to consider whether the reading passage, assessment tasks, item format, and cognitive processes remain consistent with the intended reading construct and target CEFR descriptor. Prompt C2 revises the assessment in response to instructor requirements while preserving the completed reading test specification as far as possible. When revisions involve changes to individual specification elements, such as text length or the number of assessment items, the prompt seeks to maintain the intended reading construct, target CEFR descriptor, item format, and cognitive processes established during the specification-development stage. Prompt C3 enables instructors to modify selected assessment characteristics, such as the topic, text length, or item format, while maintaining alignment with the intended reading construct and target CEFR descriptor.

The review stage is intended to provide flexibility rather than prescribe a single revision procedure. Depending on the assessment context, instructors may decide to shorten a reading passage, reduce the number of assessment items, improve distractors, adjust the level of challenge, modify the wording of individual items, or adapt the assessment to better reflect local curricular objectives. In other situations, the assessment generated during Stage 2 may already be suitable for its intended purpose and require few or no revisions. The extent of revision ultimately depends on the instructional context and the instructor's professional judgment.

Appendix K illustrates this process by presenting a revised version of the B2 assessment generated through the specification-driven approach in Stage 2 (Appendix I). In this example, the reading passage was shortened and the number of multiple-matching items was reduced in response to a classroom requirement. These revisions modified individual assessment characteristics while preserving the multiple-text format, the intended reading construct, the target CEFR descriptor, and the cognitive processes represented in the original assessment. The accompanying revision summary documents the changes that were made, explains the reasons for those changes, and describes how alignment with the completed reading test specification was maintained. The complete set of review and revision prompts is provided in Appendix C for instructors who wish to adapt the process to their own assessment development needs.

Baseline Prompting Approach

A distinguishing feature of the proposed specification-driven approach is the inclusion of an explicit specification-development stage before AI is used to generate assessment materials. As discussed in Section 3, this stage translates broad CEFR descriptors into explicit assessment characteristics that guide subsequent assessment development. To illustrate the role of this specification-development stage in the proposed approach, a baseline prompting procedure was developed to show how reading assessment generation proceeds when these assessment decisions are not first organized into a reading test specification.

Recent studies have demonstrated that AI can generate reading texts targeted at different CEFR proficiency levels using relatively direct prompts. For example, Uchida (2025) found that ChatGPT was generally able to produce texts aligned with the intended CEFR levels, while Nguyen et al. (2026) reported that AI-generated reading texts broadly reflected the expected progression of lexical demands across CEFR levels. These findings provide encouraging evidence that AI can support the development of CEFR-related reading materials. Reading assessment development, however, extends beyond generating a text at an appropriate proficiency level. It also requires explicit decisions about the reading construct, intended cognitive processes, text characteristics, item format, and scoring so that the assessment reflects what learners are expected to demonstrate. In the proposed approach, these decisions are established through the reading test specification before assessment generation begins.

This rationale informed the development of the baseline prompting procedure. To maintain the same instructional starting point as the specification-driven approach, the baseline prompt was developed from the same CEFR-aligned textbook chapters, target CEFR descriptors, and textbook reading tasks. It also included general guidance regarding the topic, target proficiency level, and reading skills to be assessed. Unlike the specification-driven approach, however, these assessment characteristics were incorporated directly into a single assessment-generation prompt rather than first being translated into an explicit reading test specification.

Both the baseline and specification-driven procedures were applied to the same B1, B1+, and B2 textbook chapters. Keeping the source materials constant allows the illustrative applications to differ only in the way assessment information is communicated to ChatGPT, making it possible to illustrate the contribution of the specification-development stage without introducing additional differences in the instructional materials or intended learning outcomes. The baseline applications are presented in Appendices F, H, and J, together with the corresponding specification-driven applications

in Appendices E, G, and I. The baseline prompt used to generate these applications is reproduced in Appendix D.

Illustrative Applications

To illustrate the proposed specification-driven approach, this section presents three illustrative applications based on CEFR-aligned textbook chapters representing the B1, B1+, and B2 proficiency levels. The applications are drawn from commercially published CEFR-aligned English language textbooks available in Thailand. The B1 application was developed from *Prepare Level 4* (Styring & Tims, 2019), the B1+ application from *Navigate Intermediate B1+* (Roberts et al., 2015), and the B2 application from *Prepare Level 7* (Styring, Tims, & Chilton, 2019). For each textbook chapter, a corresponding CEFR reading descriptor was selected because it reflected the primary reading objective of the instructional materials. These textbook chapters and descriptors served as the common instructional foundation for both the specification-driven and baseline approaches. These illustrative applications are intended to demonstrate how the proposed process can be applied in practice rather than to provide an empirical comparison of the two prompting approaches or evidence of the superiority of one approach over the other.

The specification-driven applications (Appendices E, G, and I) and the baseline applications (Appendices F, H, and J) illustrate two ways of developing CEFR-aligned reading assessments from the same instructional materials. Because both approaches began with the same textbook chapters and selected CEFR descriptors, the resulting assessment tasks shared many common characteristics. The principal difference lies in the assessment development process. In the specification-driven approach, the intended reading construct, text characteristics, item format, scoring method, and cognitive processes were documented through a reading test specification before assessment generation. In the baseline approach, these assessment considerations were incorporated directly into the assessment-generation prompt through the textbook task and CEFR descriptor without first developing a separate specification.

Across the illustrative applications, both the specification-driven and baseline approaches generated assessment tasks that broadly reflected the intended CEFR descriptors because they were grounded in the same instructional materials. A distinguishing feature of the specification-driven approach is that it makes assessment decisions explicit before assessment generation begins. By documenting the intended reading construct, cognitive processes, text characteristics, and item design through a reading test specification, the approach provides instructors with a transparent basis for

generating, reviewing, and revising AI-generated assessment materials.

This pattern can first be seen in the B1 examples (Appendices E and F). Both approaches required learners to read three short personal profiles before completing a multiple-matching task, reflecting the selected B1 descriptor that requires learners to gather information from different parts of one or more texts to complete a task. In both assessments, learners therefore needed to locate information across several profiles rather than retrieve information from a single passage.

Looking more closely at the questions, however, reveals some differences. The specification-driven assessment sampled a wider range of reading behaviours. In addition to locating explicitly stated information, learners compared information across profiles, identified future plans, and completed matching, multiple-choice, information-location, and short-answer items. For example, Item 8 required learners to identify what Ethan and Sofia had in common, requiring information to be compared across two profiles rather than retrieved from a single text. By comparison, the baseline assessment focused primarily on locating explicitly stated information, such as identifying who was preparing for a performance, who planned to attend a summer course, or who hoped to participate in a sports competition. Both assessments therefore reflected the selected B1 descriptor, but the specification-driven assessment represented a broader sample of the cognitive processes identified during the specification-development stage.

A similar pattern emerged at the B1+ level. Both approaches generated a continuous factual passage followed by comprehension questions, consistent with the selected descriptor that emphasizes understanding straightforward factual texts with satisfactory comprehension. Consequently, both assessments required learners to identify explicitly stated information and demonstrate overall comprehension of the passage.

Again, differences became more apparent when the individual questions were examined. The baseline assessment primarily targeted explicit comprehension through questions about factual information, stated reasons, viewpoints, and the overall message of the passage. The specification-driven assessment incorporated these same processes while also asking learners to identify purposes, recognize cause-and-effect relationships, make text-based inferences, and synthesize information across the passage. For example, learners explained why Daniel became interested in community service (Item 5), inferred his opinion about online friendships (Item 6), and identified the main idea of the article (Item 9). These additional question types represented a broader range of cognitive processes than simple factual retrieval and aligned closely with the processes identified during the specification-development stage.

The B2 examples provide the clearest illustration of this point. Both approaches retained the multiple-matching format found in the source textbook and therefore required learners to scan efficiently across four biographical profiles. This is consistent with the selected B2 descriptor, which emphasizes scanning quickly through longer and more complex texts to locate relevant details. Because the task format remained essentially the same, differences are more appropriately considered in terms of the information sampled by the questions rather than the task type itself.

Examining the questions more closely shows that the specification-driven assessment sampled a broader range of construct elements identified during specification development. For example, learners identified the individual who developed a project after recognizing a community problem (Item 1), intended to establish a larger organization (Item 4), hoped to reduce educational inequality (Item 7), transformed a personal interest into a successful career (Item 9), and collaborated with universities (Item 11). Although each question required learners to retrieve information from the biographies, successful performance also depended on distinguishing among similar experiences, motivations, achievements, future intentions, and organizational goals presented across four different texts.

The baseline assessment also reflected the selected B2 descriptor because learners were likewise required to scan across multiple biographies to locate relevant information. Many of the questions, however, focused on identifying a single explicitly stated detail, such as who received media attention, learned a complex skill at a young age, had not yet decided on a future career, or completed additional training. These questions remained consistent with the selected descriptor because learners still needed to locate information efficiently across multiple texts. The specification-driven assessment, by comparison, incorporated questions that represented a wider range of the cognitive processes identified during the specification-development stage, including distinguishing among similar experiences, motivations, achievements, future intentions, and organizational goals presented across four different texts.

The comparison between the two B2 applications also helps interpret the Text Inspector analyses. Across both approaches, the B1 and B1+ passages were classified at approximately their intended CEFR levels. For the B2 examples, however, the specification-driven passage was classified at C2, whereas the baseline passage was classified at C1+. At first glance, these results suggest that the linguistic characteristics of both passages exceeded the intended B2 level. When the accompanying assessment tasks are considered alongside these analyses, however, a more nuanced interpretation emerges. Despite the higher text-level classifications, both sets of questions continued to operationalize the selected B2 descriptor by requiring learners

to scan across multiple texts, locate relevant information efficiently, and distinguish among similar details. This suggests that judgments about CEFR alignment should not rely solely on automated text-level classifications but should also consider whether the assessment tasks represent the intended reading construct and cognitive processes.

The vocabulary profiles provide additional evidence supporting this interpretation. Although the specification-driven B2 passage received an overall classification of C2, fewer than five percent of the words were classified at the C1 or C2 levels by the English Vocabulary Profile. Approximately eight percent of the vocabulary consisted of words that were not assigned CEFR levels, including *workshops* and *participants*. A similar pattern was observed in the baseline application, which received an overall classification of C1+ while also containing approximately eight percent unlisted vocabulary, including *innovation* and *volunteered*. These lexical items were appropriate for the educational contexts represented in the passages, suggesting that automated CEFR estimates and vocabulary profiling capture different aspects of text difficulty and should therefore be interpreted together rather than independently. More importantly, neither analysis evaluates whether the assessment tasks themselves adequately represent the intended reading construct or cognitive processes. Those judgments continue to depend on instructor review informed by the assessment purpose, target learners, and instructional context.

Taken together, the illustrative applications demonstrate that both the specification-driven and baseline approaches can generate assessment tasks that broadly reflect the intended CEFR descriptors when they are grounded in appropriate instructional materials. Rather than relying on the textbook task and CEFR descriptor alone, the specification records the intended construct, cognitive processes, text characteristics, and item design that guide subsequent assessment development. Although the examples presented in this paper are illustrative rather than evaluative, they indicate that making these assessment decisions explicit provides instructors with a clearer basis for reviewing and refining AI-generated assessments while maintaining alignment between the selected CEFR descriptor, the intended reading construct, and the cognitive processes represented in the assessment tasks.

Practical Guidance for Instructors

The specification-driven approach presented in this paper is intended to support rather than replace instructor judgment throughout the assessment-development process. Ultimately, the quality of the assessment depends largely on the decisions instructors make before, during, and after AI generation. AI is therefore likely to be most useful when incorporated into a principled assessment-development process in which key assessment decisions are established before assessment materials are generated.

The discussion presented in Sections 2 and 3 suggests that the assessment-development process begins with a clear understanding of what the assessment is intended to measure. Because CEFR descriptors describe expected reading performance rather than assessment design, instructors must first determine how those descriptors should be translated into assessment tasks. Developing a reading test specification provides an opportunity to make these decisions explicit by identifying the intended reading construct, text type, topic domain, language expectations, item format, scoring method, and cognitive processes before any assessment materials are generated. As illustrated by the specification-driven examples in Section 4, documenting these decisions in advance provides a clear reference point for evaluating whether the generated assessment remains consistent with its intended purpose.

Once the assessment purpose has been established, instructors need an appropriate starting point for developing the specification. The examples presented in this paper suggest that CEFR-aligned textbook materials can serve this purpose because they already represent practical interpretations of CEFR descriptors within instructional contexts. Rather than asking AI to generate an assessment directly from a CEFR level label, instructors may begin by examining the reading task in a textbook chapter and identifying the reading behaviours it is designed to elicit. A reading test specification can then be developed from these observations and incorporated into AI prompts to generate a new assessment that reflects the same instructional objectives while using different reading passages and assessment tasks.

The completed specification subsequently serves as the foundation for AI-assisted assessment generation. As illustrated by the three illustrative applications, both approaches generated assessment tasks that broadly reflected the intended CEFR descriptors because both were grounded in the same textbook chapters and descriptors. A distinguishing feature of the specification-driven approach is that the intended reading construct, cognitive processes, text characteristics, and item design are documented explicitly before assessment generation begins. This provides instructors with a transparent record of the assessment decisions that guide subsequent prompting and revision. In practice, instructors may also find it helpful to identify a small number of assessment characteristics that should remain unchanged throughout the generation process, such as text length, text organization, number of items, item format, and target cognitive processes.

Generating assessment materials should not be regarded as the final stage of the process. Instead, the generated assessment should be reviewed as a whole before it is used in the classroom. The B2 examples discussed in Section 4 illustrate why this stage is important. Although the generated passages received higher overall CEFR estimates from Text Inspector than

intended, the accompanying assessment tasks continued to reflect the cognitive processes represented in the selected B2 descriptor. This observation suggests that instructors should evaluate both the linguistic characteristics of the reading passage and the cognitive demands represented by the accompanying questions. Automated tools such as Text Inspector provide useful information about vocabulary profiles, readability, and estimated CEFR levels, but these results should be interpreted together with the assessment tasks, instructional objectives, and intended reading construct.

Reviewing the assessment questions is equally important. The illustrative applications showed that questions based on similar reading passages can nevertheless differ in the cognitive processes they require. Some questions primarily require learners to locate explicitly stated information, whereas others ask them to compare information across texts, recognize relationships among ideas, distinguish among similar details, or integrate information from different parts of a passage. Reviewing questions from this perspective helps ensure that the assessment represents the intended reading construct rather than emphasizing only the retrieval of isolated pieces of information.

The same principle applies to response options in multiple-choice items. Distractors should be plausible, supported by information contained in the passage, and comparable in length and structure to the correct answer. More importantly, they should require learners to engage in the intended reading processes rather than rely on superficial word matching or the elimination of obviously incorrect alternatives. Because AI-generated materials can be revised iteratively, instructors can refine passages, questions, and response options until they more closely reflect the intended assessment purpose, learner characteristics, and instructional context.

The examples presented throughout this paper suggest that AI can make assessment development more efficient, but they also reinforce the continuing importance of instructor judgment. Decisions about construct representation, cognitive processes, task appropriateness, and assessment quality cannot be delegated entirely to AI or to automated text analysis. Instead, they remain the responsibility of instructors, who are best positioned to evaluate whether an assessment is appropriate for their learners and consistent with their instructional goals. Used in this way, AI becomes a tool that supports principled assessment development rather than a substitute for assessment expertise.

Implications and Conclusion

This paper has presented a specification-driven approach to AI-assisted reading assessment development in which reading test specifications

derived from CEFR-aligned instructional materials were incorporated into AI prompts to guide the generation of reading assessments. The approach was illustrated through three illustrative applications at the B1, B1+, and B2 levels and was also compared with a baseline prompting approach developed from the same textbook chapters, reading tasks, and CEFR descriptors. These illustrative applications were intended to demonstrate how the proposed specification-driven approach can be implemented in practice rather than to provide an empirical evaluation of its effectiveness or establish the superiority of one prompting approach over another.

The illustrative applications suggest that both the specification-driven and baseline approaches were capable of generating assessment tasks that broadly reflected the intended CEFR descriptors when they were grounded in the same instructional materials. At the same time, the examples indicate that the principal contribution of the specification-driven approach lies in making assessment decisions explicit before assessment generation begins. By documenting the intended reading construct, cognitive processes, text characteristics, and item design through a reading test specification, the approach provides a transparent basis for generating, reviewing, and revising AI-generated assessment materials.

These examples also highlight the important role of CEFR-aligned instructional materials in the assessment-development process. In the present study, the reading test specifications were not developed directly from CEFR descriptors alone but were derived from textbook chapters and accompanying reading tasks that had already been designed to reflect the intended proficiency levels. This provided a concrete instructional context for interpreting broad CEFR descriptors and identifying the reading behaviours, task demands, and assessment characteristics to be represented in the generated assessments. The examples therefore suggest that well-designed instructional materials, including commercially published CEFR-aligned textbooks and carefully developed classroom handouts, can provide valuable starting points for AI-assisted assessment development. Beginning with instructional materials rather than a CEFR level label alone also helps maintain alignment between instruction and assessment because the resulting assessment is derived from the same learning objectives and reading behaviours that learners have encountered during instruction.

More broadly, this approach may help strengthen the connection between classroom instruction and assessment. In many instructional contexts, teachers invest considerable time in selecting or adapting CEFR-aligned instructional materials and designing learning activities to achieve specific learning objectives. By deriving reading test specifications from those same instructional materials, AI-assisted assessment development has the potential to support the creation of classroom assessments that reflect the

learning objectives, reading behaviors, and cognitive processes emphasized during instruction. Although further research is needed to examine its implementation in different educational contexts, this approach may provide a practical means of promoting closer alignment between what is taught and what is assessed.

The B2 examples further demonstrate why AI-generated assessments should be evaluated as complete assessment tasks rather than as reading passages alone. Although Text Inspector classified the generated passages above the intended CEFR level, the accompanying assessment tasks continued to reflect the cognitive processes represented in the selected B2 descriptor. Similarly, the vocabulary profile analyses showed that overall CEFR estimates and vocabulary profiling provided different but complementary sources of evidence. Together, these observations suggest that automated analyses should be interpreted alongside the assessment tasks, the intended reading construct, and instructor judgment when evaluating CEFR alignment.

The findings presented in this paper should, however, be interpreted within the scope of the illustrative applications. The examples demonstrate an assessment-development procedure rather than the effectiveness of the resulting assessments. Although the specification-driven examples appeared to represent a broader range of construct elements identified during specification development, the present paper does not establish that one prompting approach produces better assessments than the other. Nor does it provide evidence regarding reliability, fairness, scoring consistency, item difficulty, item discrimination, distractor effectiveness, or construct validity. These aspects of assessment quality require empirical investigation with learners.

The illustrative applications nevertheless provide practical implications for instructors who wish to incorporate AI into assessment development. Rather than beginning with a CEFR level label alone, instructors may find it beneficial to begin with instructional materials that are already aligned with the intended CEFR level, develop a reading test specification that makes assessment decisions explicit, and then use that specification to guide AI-assisted assessment generation. Throughout this process, instructor judgment remains essential for evaluating the relationship among the reading passage, assessment tasks, cognitive processes, and intended reading construct. AI may assist by producing initial drafts of reading passages and assessment tasks more efficiently, but careful review and revision by instructors remain necessary to ensure that the assessments accurately represent the intended construct and are appropriate for the local instructional context.

Several limitations of the present paper should also be acknowledged. First, the proposed approach was illustrated only through reading assessment and therefore may not transfer directly to listening, speaking, or writing assessment without further investigation. Second, the examples were intended to demonstrate the assessment-development process rather than compare prompting approaches systematically. Future research should empirically examine the implementation of the proposed approach in classroom settings, investigate its effects on assessment quality and learner outcomes, compare its application across different educational contexts and CEFR levels, and evaluate its use with other large language models. Such research should also examine validity evidence, reliability, fairness, item quality, and standard-setting procedures to determine the extent to which AI-assisted assessments developed through this approach are appropriate for operational use.

In conclusion, this paper has presented one approach to incorporating AI into CEFR-aligned reading assessment development while maintaining established principles of language assessment. Rather than evaluating the effectiveness of AI prompting approaches, the paper has illustrated how a specification-driven process can support the development of CEFR-aligned classroom reading assessments through three illustrative applications. The proposed approach emphasizes the importance of deriving reading test specifications from CEFR-aligned instructional materials and making assessment decisions explicit before AI generation. Within this process, AI functions as a tool that supports assessment development, while responsibility for assessment design, evaluation, and refinement remains firmly with the instructor. Used in this way, AI has the potential to support more transparent, systematic, and instructionally aligned development of CEFR-based classroom reading assessments.

About the Authors

Sun-Young Shin: a Professor in the Department of Linguistics at Indiana University. His research interests include GenAI-mediated language assessment, L2 listening assessment, L2 pragmatics assessment, and standard setting. His work has been published in numerous journals and edited volumes, and he has been invited to deliver lectures and workshops on L2 assessment in several countries. He currently serves on the editorial boards of *Language Testing* and *Language Assessment Quarterly*.

Suchada Sanonguthai: A Lecturer at the Language Institute, Thammasat University. She holds a Ph.D. in Second Language Studies from Indiana University, with a specialization in Second Language Assessment and a minor

in Inquiry Methodology. Her research focuses on test validation, CEFR-aligned test development, standard setting, and differential item functioning.

References

- Aryadoust, V., & Wong, J. (2026). How to train your dragon: Evaluating prompting and fine-tuning for GPT-based item generation in L2 listening assessment. *Computers and Education: Artificial Intelligence*, 10, 100623. <https://doi.org/10.1016/j.caeai.2026.100623>
- Bax, S. (n.d.). *Text Inspector*. Retrieved July 5, 2026, from <https://textinspector.com/>
- Charttrakul, K., & Damnet, A. (2021). Role of the CEFR and English teaching in Thailand: A case study of Rajabhat Universities. *Advances in Language and Literary Studies*, 12(2), 82–89. <https://journals.aiac.org.au/index.php/all/article/view/6666>
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/1680459f97>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Roberts, R., Buchanan, H., & Pathare, E. (2015). *Navigate intermediate B1+ coursebook with video and Oxford online skills*. Oxford University Press.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/Information-on-the-BNC-COCA-word-family-lists.pdf>
- Nguyen, N. T., Huang, X., & Dang, T. N. Y. (2026). Lexical profile of ChatGPT-generated reading materials targeting EFL learners across the CEFR levels. *International Journal of TESOL Studies*, 8(3), 113–133. <https://doi.org/10.58304/ijts.260211>
- OpenAI. (n.d.). *ChatGPT (GPT-5.5)* [Large language model]. Retrieved June 30, 2026, from <https://chatgpt.com>
- Phoolaikao, W., & Sukying, A. (2021). Insights into CEFR and its implementation through the lens of preservice English teachers in Thailand. *English Language Teaching*, 14(6), 25–35. <https://doi.org/10.5539/elt.v14n6p25>
- Piamsai, C. (2023). Development and use of CEFR-based self-assessment in

- English language learning and assessment in Thailand. *PASAA*, 66, 45–76. <https://doi.org/10.58837/CHULA.PASAA.66.1.3>
- Shin, S.-Y. (2026). Proficiency scales. In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics* (2nd ed.). Wiley. <https://doi.org/10.1002/9781405198431.wbeal1423.pub2>
- Siripol, P., Rhee, S., Thirakunkovit, S., & Liang-Itsara, A. (2025). Evaluating the consistency of automated CEFR analyzers: a study of English language text classification. *International Journal of Evaluation and Research in Education*, 14(4), 3283–3294. <https://doi.org/10.11591/ijere.v14i4.33528>
- Styring, J., & Tims, N. (2019). *Prepare level 4 student's book* (2nd ed.) Cambridge University Press.
- Styring, J., Tims, N., & Chilton, H. (2019). *Prepare level 7 student's book* (2nd ed.). Cambridge University Press.
- Uchida, S. (2025). Generative AI and CEFR levels: Evaluating the accuracy of text generation with ChatGPT-4o through textual features. *Vocabulary Learning and Instruction*, 14(1), 1–13. <https://doi.org/10.29140/vli.v14n1.2078>
- Waluyo, B. (2019). Thai first-year university students' English proficiency on CEFR levels: A case study of Walailak University, Thailand. *The New English Teacher*, 13(2), 51–71. <http://www.assumptionjournal.au.edu/index.php/newEnglishTeacher/article/view/3651>
- Wudthayagorn, J. (2022). An exploration of the English exit examination policy in Thai public universities. *Language Assessment Quarterly*, 19(2), 107–123. <https://doi.org/10.1080/15434303.2021.1937174>

Appendix A

Prompt A: Reading Test Specification Development

Textbook Chapter to be Analyzed

[Upload the textbook chapter]

You are an experienced language assessment specialist with expertise in CEFR-aligned reading assessment design.

Your task is to analyze a CEFR-aligned textbook chapter and develop a reading test specification based on the chapter's reading materials and the target CEFR descriptor.

Focus only on reading comprehension. Do not include speaking, listening, writing, pronunciation, grammar, or vocabulary as separate assessment objectives. Consider vocabulary and grammar only when they help students understand the reading texts.

The following CEFR descriptor has already been reviewed and agreed upon by teachers as the target learning outcome. Use this descriptor as the main guide when making decisions about the test specification:

[Insert the CEFR descriptor]

What to Analyze

Look at the whole chapter rather than a single activity.

Use the following as evidence:

- Reading texts
- Reading comprehension activities
- Reading-related learning objectives
- Topics and themes found in the reading materials
- Vocabulary and grammar features that help students understand the reading texts

Do not use the following as evidence unless they directly support reading comprehension:

- Speaking activities
- Listening activities
- Writing activities
- Pronunciation activities
- Grammar practice activities
- Vocabulary practice activities
- Other language production activities

If grammar or vocabulary sections appear in the chapter, include them only when they support understanding of the reading texts.

Principles for Developing the Specification

The specification should:

- Be based only on reading-related content in the chapter.
- Reflect the target CEFR descriptor.
- Represent the chapter as a whole.
- Use evidence from reading texts, reading activities, learning goals, topics, and language features that support comprehension.
- Provide enough detail to guide the development of a new reading assessment.

- Give more weight to reading comprehension activities because they show the reading skills students are expected to use.
- Be suitable for creating a new parallel reading assessment rather than copying the original textbook task.
- When determining text length, use the chapter's reading text(s) as the primary source of evidence. Recommendations should reflect the approximate length and structure of the chapter's reading materials rather than generic expectations associated with the CEFR level.

Specification Components

Complete all components in the table below.

Specification Component	Description
Target CEFR Level	The intended proficiency level of the assessment
Reading Construct	The reading ability or abilities being measured
Text Type	The type of text to be used (e.g., article, email, profile, biography, blog post)
Topic Domain	The topic area or context of the reading text
Text Length	Approximate length of the reading text, based on the chapter's reading materials. If the chapter contains multiple texts or sections, indicate the approximate total length and the number of texts or sections.
Vocabulary and Grammar Expectations	Language features that support comprehension at the target level
Item Format	Format of assessment items (e.g., multiple choice, matching, short answer)
Number of Items	Total number of assessment items
Intended Cognitive Process	The reading processes needed to answer the items
Scoring Method	How responses will be scored

Guiding Questions

Consider the following questions when developing the specification:

1. What CEFR level is represented by the descriptor and the chapter's reading materials?
2. What reading skills are students expected to use?
3. What text types are used in the chapter's reading materials?
4. What topics and themes appear in the reading materials?
5. What text length is appropriate for this CEFR level?
6. What vocabulary and grammar features support comprehension at this level?
7. What item formats best reflect the reading skills being assessed?
8. How many items are needed to provide enough evidence of reading ability?
9. What reading processes do students need to complete the task successfully?
10. How should responses be scored?

Output Requirements

Generate the completed reading test specification in table format.

For each specification component, provide:

1. Recommended Specification
2. Rationale

The rationale should:

- Refer to the target CEFR descriptor.
- Refer to evidence from the chapter's reading texts and reading comprehension activities.
- Explain how the recommendation supports the intended reading construct.
- For Text Length, explain how the recommendation reflects the approximate length and structure of the chapter's reading text(s), including the number of texts or sections when relevant.

For Vocabulary and Grammar Expectations, specify:

- Appropriate vocabulary range for the target CEFR level
- Expected degree of lexical diversity
- Whether key vocabulary should be recycled throughout the text
- Expected sentence complexity
- Any vocabulary or grammar features that support comprehension of the reading texts

Whenever possible, base recommendations on evidence from several reading-related features of the chapter rather than a single text or activity.

Summary of Key Assessment Characteristics

After generating the completed reading test specification, provide a final section titled Summary of Key Assessment Characteristics.

Summarize the completed reading test specification using bullet points under the following headings:

- Target CEFR Level
- Reading Construct
- Text Type
- Topic Domain
- Text Length
- Vocabulary and Grammar Expectations
- Item Format
- Number of Items
- Intended Cognitive Process
- Scoring Method

The summary should provide a concise overview of the completed reading test specification. It should be fully consistent with the specification table and should not introduce any additional assessment characteristics.

This summary will serve as the primary reference for the reading assessment generation stage.

Appendix B

Prompt 2: Reading Assessment Generation

[Insert the completed reading test specification and the summary of key assessment characteristics generated in Stage 1 here]

You are an experienced language assessment specialist with expertise in CEFR-aligned reading assessment design.

The reading test specification above serves as the blueprint for the assessment. Your task is to create a new reading assessment that follows the specification as closely as possible.

Assessment Development Requirements
Generate:

1. A reading passage (or set of texts, if specified in the reading test specification).
2. Assessment items that match the specified item format.
3. An answer key with supporting evidence from the text.

When developing the assessment:

- Follow all components of the reading test specification.
- Ensure that the reading passage reflects the target CEFR level and uses vocabulary and grammar consistent with the specification.
- Use vocabulary, grammatical structures, and overall readability that are consistent with the linguistic expectations specified in the reading test specification and typical of authentic CEFR-aligned textbook readings at the target level.
- Create entirely new content while preserving the reading construct, text type, and intended cognitive processes specified in the test specification.
- Ensure that all information required to answer the questions is provided in the text.
- Ensure that all assessment items measure the intended reading construct and cognitive processes.
- Ensure that the assessment provides sufficient opportunities for learners to demonstrate each intended cognitive process identified in the reading test specification.
- Ensure that every question can be answered using information from the text and that only one correct answer is possible.
- Avoid trick questions, ambiguous items, and questions that rely on background knowledge, personal experience, or opinion.

Output Format

Reading Passage

[Insert reading passage(s)]

Assessment Items

[Insert assessment items]

Answer Key and Evidence

For each item, provide:

- The correct answer.
- The location of the supporting information in the text.
- The reading skill or cognitive process being assessed.
- A brief explanation of why the answer is correct.

Quality Check

Before presenting the final assessment, verify that:

1. The assessment matches the reading test specification.
2. The linguistic difficulty and readability are appropriate for the target CEFR level.
3. The item format, number of items, and cognitive processes match the specification.
4. The assessment provides evidence of each intended cognitive process identified in the specification.
5. The content is new and consistent with the specification.
6. All questions can be answered using information from the text.
7. Only one correct answer is possible for each item.
8. Supporting evidence has been provided for every answer.
9. The assessment measures the intended reading construct and only reading comprehension.

Generate the completed reading assessment only. Do not provide commentary on the development process.

Appendix C

Prompt C1: Revision/Checking Prompt for CEFR Alignment and Item Quality

Reading Test Specification

[Insert the completed reading test specification and the summary of key assessment characteristics generated in Stage 1 here]

Generated Reading Assessment

[Insert the reading assessment generated in Stage 2 here]

You are an experienced language assessment specialist with expertise in CEFR-aligned reading assessment design.

The reading test specification serves as the benchmark for evaluation. Review the generated reading assessment and determine whether it faithfully implements the specification and target CEFR level.

CEFR Alignment

Evaluate whether:

- The linguistic difficulty is appropriate for the target CEFR level.
- The vocabulary profile is consistent with the Vocabulary and Grammar Expectations specified in the test specification.
- The sentence complexity is appropriate for the target CEFR level.
- The lexical diversity is appropriate for the target CEFR level.
- The overall reading demands are consistent with the target CEFR descriptor.

Specification Alignment

Evaluate whether:

- The assessment follows the specified text type.
- The assessment follows the specified topic domain.
- The assessment follows the specified text length.
- The assessment uses the specified item format.
- The assessment contains the specified number of items.
- The assessment measures the specified reading construct.
- The assessment requires the intended cognitive processes identified in the specification.

Item Quality

Review all assessment items and determine whether:

- Each item has one clearly correct answer.
- All distractors are plausible but clearly incorrect.
- Multiple-choice options are balanced in length and level of detail.
- No grammatical, lexical, or stylistic clues reveal the correct answer.
- No item depends primarily on background knowledge, personal experience, or opinion.
- No item can be answered without reading the text.
- No item is ambiguous.
- No item is a trick question.
- The answer key is accurate and supported by evidence from the text.

Output Format

CEFR Alignment Review

Specification Alignment Review

Item Quality Review

Recommended Revisions

For each issue identified:

- Describe the problem.
- Explain why it affects CEFR alignment, construct representation, or item quality.
- Provide a specific revision.

Overall Judgment

Choose one:

- Accept without revision
- Accept with minor revision
- Accept with major revision
- Regenerate assessment

Do not generate a completely new assessment unless regeneration is recommended.

Prompt C2: Revision Prompt for Classroom Constraints

Reading Test Specification

[Insert the completed reading test specification and the summary of key assessment characteristics generated in Stage 1 here]

Generated Reading Assessment

[Insert the reading assessment generated in Stage 2 here]

Additional Classroom Requirements

[Insert classroom constraints, e.g., shorter text, fewer items, shorter testing time]

You are an experienced language assessment specialist with expertise in CEFR-aligned reading assessment design.

Revise the assessment to meet the additional classroom requirements while maintaining alignment with the reading test specification and target CEFR level.

Revision Requirements

- Preserve the original reading construct.
- Preserve the target CEFR level.
- Preserve the intended cognitive processes as specified in the reading test specification.
- Preserve the topic domain and communicative purpose unless otherwise specified.
- Ensure that the revised assessment remains valid for measuring the

intended reading ability.

- If the text length or number of items must be reduced, retain the most important evidence of reading comprehension.
- Ensure that the revised assessment remains appropriate for the specified classroom context.

Output Format

Revised Reading Passage

[Insert revised passage]

Revised Assessment Items

[Insert revised items]

Revision Summary

Briefly explain:

- What changes were made.
- Why the changes were necessary.
- How CEFR alignment was maintained.
- How the intended cognitive processes were preserved.

Prompt C3: Revision Prompt for Vocabulary and Language Targets

Reading Test Specification

[Insert the completed reading test specification and the summary of key assessment characteristics generated in Stage 1 here]

Generated Reading Assessment

[Insert the reading assessment generated in Stage 2 here]

Additional Language Targets

[Insert target vocabulary, grammar features, or language-learning objectives here]

You are an experienced language assessment specialist with expertise in CEFR-aligned reading assessment design.

Revise the assessment so that it incorporates the specified language targets while maintaining alignment with the reading test specification and target CEFR level.

Revision Requirements

- Preserve the original reading construct.

- Preserve the target CEFR level.
- Preserve the intended cognitive processes.
- Maintain the original text type and topic domain unless otherwise specified.
- Incorporate the specified language targets naturally within the reading passage.
- Do not convert the assessment into a vocabulary or grammar test.
- Ensure that vocabulary and grammar continue to support reading comprehension rather than become the primary assessment focus.
- Ensure that item answers remain recoverable from the text.
- Ensure that the revised assessment remains valid as a reading comprehension assessment.

Output Format

Revised Reading Passage

[Insert revised passage]

Revised Assessment Items

[Insert revised items if changes are required]

Revision Summary

Briefly explain:

- What changes were made.
- Why the changes were necessary.
- How CEFR alignment was maintained.
- How the intended cognitive processes were preserved.

Appendix D

Baseline Prompt: Textbook-Informed Assessment Generation

Textbook Reading Task

[Insert textbook reading task here]

The following CEFR descriptor has already been reviewed and agreed upon by teachers as the target learning outcome:

[Insert the CEFR descriptor]

Using the textbook reading task and the CEFR descriptor as references, create a new reading comprehension assessment. Focus only on reading comprehension.

Generate:

1. A reading passage (or set of texts, if appropriate).
2. Reading comprehension questions.
3. An answer key with supporting evidence from the text.

Requirements

- Use the textbook reading task to guide the topic, level, and reading skills.
- Ensure that the assessment is appropriate for the target CEFR level.
- Reflect the communicative reading abilities described in the CEFR descriptor.
- Create a new assessment that could reasonably be used as a parallel task for learners studying the same textbook unit.
- Do not copy or closely paraphrase the original textbook text.

Output Format

Reading Passage

[Insert reading passage]

Questions

[Insert questions]

Answer Key with Supporting Evidence

For each question, provide:

- The correct answer.
- The location of the supporting information in the text.
- The reading skill or cognitive process being assessed.
- A brief explanation of why the answer is correct.

Appendix E

B1 Example of the Specification-Driven Approach: Test Specification, Assessment, and Text Inspector Results

Textbook: *Prepare Level 4* (Styring & Tims, 2019)

Publisher: Cambridge University Press

Target CEFR Level: B1

Unit: Unit 1, *All About Me*

Source Reading Task: Learners read three short online personal profiles and complete a multiple-matching task by locating explicit information, comparing information across profiles, and identifying future plans, interests, and personal characteristics.

Selected CEFR Reading Descriptor: *Can scan longer texts in order to locate desired information and gather information from different parts of a text, or from different texts, in order to fulfil a specific task* (Council of Europe, 2020, p. 55).

Reading Test Specification

Specification Component	Recommended Specification	Rationale
Target CEFR Level	B1	The target descriptor emphasizes scanning longer texts and gathering information from different parts of one or more texts to complete a task. The chapter requires learners to read three student profiles, compare information across them, and identify explicit details, future plans, interests, and personal characteristics.
Reading Construct	Ability to scan multiple short texts to locate specific information, identify explicit details, compare information across texts, and integrate information from different profiles to complete reading tasks.	The reading activities ask learners to identify which student matches specific descriptions, locate factual information, and compare information across three online profiles, directly reflecting the target descriptor.

Specification Component	Recommended Specification	Rationale
Text Type	Three short online personal profiles presented on a school website or social networking platform.	The chapter's reading input consists of three online student profiles ("Bryans High School all.about.me"), making this the most appropriate text type for a parallel assessment.
Topic Domain	Personal information, hobbies, interests, personality, family, school life, leisure activities, and future plans.	The reading texts consistently focus on familiar personal topics appropriate for B1 learners, including hobbies, family, interests, personality, competitions, and future intentions.
Text Length	Approximately 250–350 words distributed across three short profiles (approximately 80–120 words each).	The chapter contains three relatively short profiles that together form a multiple-text reading task. Maintaining a similar structure preserves the intended scanning and comparison demands.
Vocabulary and Grammar Expectations	Vocabulary should consist primarily of high-frequency B1 vocabulary related to personal identity, hobbies, family, school, technology, entertainment, sports, travel, and future plans. Lexical diversity should be moderate, with key vocabulary recycled across the three texts. Grammar should primarily include present simple, present continuous, basic future expressions, modals, and straightforward compound and simple complex sentences. Vocabulary and grammar should support comprehension rather	Although the chapter contains grammar and vocabulary sections, they reinforce comprehension of the reading texts by highlighting recurring language used in the profiles (e.g., present simple and present continuous for describing people and current activities). The lexical content remains familiar and accessible.

Specification Component	Recommended Specification	Rationale
	than increase linguistic difficulty.	
Item Format	Matching items, multiple-choice questions, information-location items, and short-answer questions assessing explicit comprehension.	The chapter's reading activities include matching interests, completing information, identifying students who match statements, and locating specific information, making these formats appropriate for a parallel assessment.
Number of Items	10–12 items	This number provides sufficient evidence of learners' ability to locate, compare, and integrate information across multiple texts while maintaining a classroom-appropriate assessment length comparable to the chapter.
Intended Cognitive Process	• Scanning for specific information • Identifying explicitly stated information • Matching information to individuals • Comparing information across texts • Integrating information from multiple texts to complete a task • Identifying future plans, interests, and personal characteristics	These processes are consistently required throughout the reading comprehension activities, particularly when learners identify which student matches different descriptions and locate supporting details across the three profiles.
Scoring Method	Dichotomous scoring (1 point for each correct response; 0 points for each incorrect response). Report total score and percentage correct.	The reading tasks have objectively correct responses and assess successful retrieval and comparison of information, making dichotomous scoring appropriate.

Summary of Key Assessment Characteristics

- Target CEFR Level
 - B1
- Reading Construct
 - Ability to scan multiple short texts to locate specific information, identify explicit details, compare information across texts, and integrate information from different profiles to complete reading tasks.
- Text Type
 - Three short online personal profiles presented on a school website or social networking platform.
- Topic Domain
 - Personal information, hobbies, interests, personality, family, school life, leisure activities, and future plans.
- Text Length
 - Approximately 250–350 words across three short profiles (about 80–120 words each).
- Vocabulary and Grammar Expectations
 - High-frequency B1 vocabulary related to familiar personal topics.
 - Moderate lexical diversity with recycling of key vocabulary across texts.
 - Primarily present simple, present continuous, basic future forms, modals, and straightforward sentence structures that support comprehension.
- Item Format
 - Matching, multiple-choice, information-location, and short-answer questions.
- Number of Items
 - 10–12 items.
- Intended Cognitive Process
 - Scanning for specific information.
 - Identifying explicitly stated information.
 - Matching information to individuals.
 - Comparing information across texts.
 - Integrating information from multiple texts.
 - Identifying future plans, interests, and personal characteristics.
- Scoring Method
 - Dichotomous scoring (1 point per correct response, 0 points per incorrect response), reported as both a raw score and percentage.

Reading Passage

Student Connect

Read the profiles of three students on a school website.

Maya

Hi! I'm Maya and I'm 15 years old. I live with my parents and my younger brother. I'm quite friendly, but I'm also a little shy when I meet new people.

My favorite hobby is photography. At weekends, I usually take photos in the city and post them online. Right now, I'm taking part in a photography competition for teenagers. If I win, I'll receive a new camera.

I'm also interested in travelling. Next summer, my family is going to visit Japan. I'm really looking forward to learning more about Japanese culture and food.

Ethan

Hello! My name's Ethan. I live with my mum and my grandparents. Many people think I'm serious, but my friends know that I enjoy making jokes.

I'm really into technology and computer games. At the moment, I'm learning to design mobile apps with help from my uncle. He works as a software engineer and teaches me new skills every weekend.

Recently, I joined a school technology club. We meet every Friday after school to work on projects together. One day, I'd like to create my own company.

Sofia

Hi everyone! I'm Sofia. I'm an only child and I live with my parents. I'm cheerful and patient, and I love spending time outdoors.

My favorite activity is sport. I play volleyball for the school team and I also enjoy cycling. This year, I'm having extra volleyball training because our team might compete in a regional tournament.

In a few months, I'm going on a cycling trip with my cousin. She is an experienced cyclist and has been helping me prepare for the trip. I hope the experience will help me become a stronger athlete.

(288 words)

Student Connect

Read the profiles of three students on a school website.

Maya

Hi! I'm Maya and I'm 15 years old. I live with my parents and my younger brother. I'm quite friendly, but I'm also a little shy when I meet new people.

My favorite hobby is photography. At weekends, I usually take photos in the city and post them online. Right now, I'm taking part in a photography competition for teenagers. If I win, I'll receive a new camera.

I'm also interested in travelling. Next summer, my family is going to visit Japan. I'm really looking forward to learning more about Japanese culture and food.

Ethan

Hello! My name's Ethan. I live with my mum and my grandparents. Many people think I'm serious, but my friends know that I enjoy making jokes.

I'm really into technology and computer games. At the moment, I'm learning to design mobile apps with help from my uncle. He works as a software engineer and teaches me new skills every weekend.

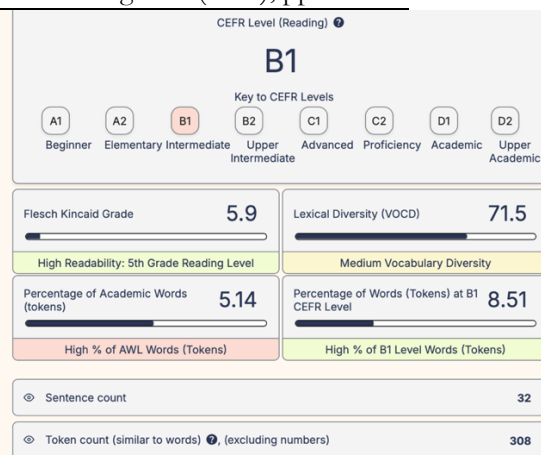
Recently, I joined a school technology club. We meet every Friday after school to work on projects together. One day, I'd like to create my own company.

Sofia

Hi everyone! I'm Sofia. I'm an only child and I live with my parents. I'm cheerful and patient, and I love spending time outdoors.

My favorite activity is sport. I play volleyball for the school team and I also enjoy cycling. This year, I'm having extra volleyball training because our team might compete in a regional tournament.

In a few months, I'm going on a cycling trip with my cousin. She is an experienced cyclist and has been helping me prepare for the trip. I hope the experience will help me become a stronger athlete.



Assessment Items

Part A: Matching (Questions 1–4)

Match each statement with the correct student.

Write:

A = Maya

B = Ethan

C = Sofia

- _____ is learning a new skill from a family member.
- _____ may receive a prize soon.
- _____ is preparing for a sports competition.
- _____ wants to start a business in the future.

Part B: Multiple-Choice Questions (Questions 5–8)

Choose the best answer.

- Why is Maya excited about next summer?
 - She is moving to another country.
 - She is planning to learn about a different culture.
 - She is starting a new school.
 - She is meeting her cousin.
- What does Ethan do every Friday?
 - Visits his uncle
 - Plays computer games
 - Attends a technology club
 - Studies with his grandparents

7. Which statement is true about Sofia?
 - A. She has a younger brother.
 - B. She dislikes outdoor activities.
 - C. She belongs to a school sports team.
 - D. She lives with her cousin.

8. What do Ethan and Sofia have in common?
 - A. Both are only children.
 - B. Both are learning from relatives.
 - C. Both are entering photography competitions.
 - D. Both are travelling abroad soon.

Part C: Information Location (Questions 9–10)

Write ONE WORD OR A SHORT PHRASE.

9. What hobby does Maya participate in competitively?

10. What kind of apps is Ethan learning to design?

Part D: Short Answer (Questions 11–12)

Answer in no more than five words.

11. Why is Sofia having extra volleyball training?

12. Who helps Ethan learn app design?

Answer Key and Evidence

Item	Answer	Supporting Information	Reading Skill/ Cognitive Process	Explanation
1	B (Ethan)	“I’m learning to design mobile apps with help from my uncle.”	Identifying explicit details	Ethan is learning a new skill from a relative.
2	A (Maya)	“If I win, I’ll receive a new camera.”	Scanning for specific information	Maya may receive a prize from the competition.

Item	Answer	Supporting Information	Reading Skill/ Cognitive Process	Explanation
3	C (Sofia)	“Our team might compete in a regional tournament.”	Identifying explicit details	Sofia is preparing for a possible sports competition.
4	B (Ethan)	“One day, I’d like to create my own company.”	Locating future plans	Ethan wants to start a business.
5	B	“My family is going to visit Japan. I’m really looking forward to learning more about Japanese culture and food.”	Identifying explicit details	Maya is excited about learning about another culture.
6	C	“We meet every Friday after school to work on projects together.”	Scanning for specific information	The technology club meets every Friday.
7	C	“I play volleyball for the school team.”	Identifying explicit details	Sofia belongs to a school sports team.
8	B	Ethan learns from his uncle; Sofia is going on a cycling trip with her cousin.	Comparing information across texts	Both profiles involve activities connected to relatives.
9	Photography	“I’m taking part in a photography competition.”	Information retrieval	The competition is related to photography.
10	Mobile apps	“I’m learning to design mobile apps.”	Information retrieval	Ethan is learning app design.

Item	Answer	Supporting Information	Reading Skill/ Cognitive Process	Explanation
11	To compete in a tournament	“Our team might compete in a regional tournament.”	Locating supporting details	Extra training is connected to the tournament.
12	His uncle	“With help from my uncle.”	Information retrieval	Ethan’s uncle teaches him app design.

Appendix F

B1 Example of the Baseline Approach: Assessment and Text Inspector Results

Textbook: *Prepare Level 4* (Styring & Tims, 2019)

Publisher: Cambridge University Press

Target CEFR Level: B1

Unit: Unit 1, *All About Me*

Source Reading Task: Learners read three short online personal profiles and complete a multiple-matching task by locating explicit information, comparing information across profiles, and identifying future plans, interests, and personal characteristics.

Selected CEFR Reading Descriptor: *Can scan longer texts in order to locate desired information and gather information from different parts of a text, or from different texts, in order to fulfil a specific task* (Council of Europe, 2020, p. 55).

Reading Passage

MyStudentSpace

MyStudentSpace is a website where teenagers can meet other students who share similar interests. Read the profiles below.

Jake

Hi everyone! I’m Jake and I’m 15 years old. I live with my parents and my younger sister. I’m interested in technology and computer games. After school, I often spend time learning how to create websites.

Right now, I'm taking part in a competition for young programmers. My older cousin is teaching me some new computer skills. In the future, I'd like to work in software design. I enjoy solving problems and creating new things.

Mia

Hello! My name is Mia. I'm 15 and I live with my mother and father. My favorite hobby is music. I play the guitar and sing in a school band.

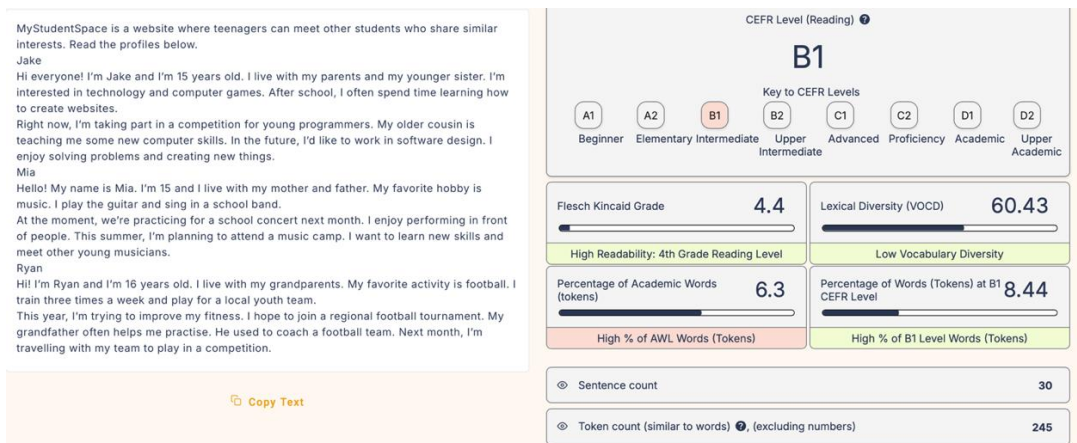
At the moment, we're practicing for a school concert next month. I enjoy performing in front of people. This summer, I'm planning to attend a music camp. I want to learn new skills and meet other young musicians.

Ryan

Hi! I'm Ryan and I'm 16 years old. I live with my grandparents. My favorite activity is football. I train three times a week and play for a local youth team.

This year, I'm trying to improve my fitness. I hope to join a regional football tournament. My grandfather often helps me practise. He used to coach a football team. Next month, I'm travelling with my team to play in a competition.

(240 words)



Questions

Part A: Match the Person to the Information

Write:

A = Jake

B = Mia

C = Ryan

1. Who receives help from a relative?
2. Who is preparing for a performance?
3. Who wants a career related to technology?
4. Who plans to attend a special course during the summer?
5. Who hopes to participate in a sports competition?

Part B: Multiple Choice

6. Why is Jake taking part in a competition?
 - A. To improve his football skills
 - B. To learn about music
 - C. To develop computer skills
 - D. To meet musicians
7. What is Mia preparing for?
 - A. A music camp
 - B. A school concert
 - C. A football tournament
 - D. A family trip
8. Why does Ryan practise with his grandfather?
 - A. His grandfather plays in the same team.
 - B. His grandfather teaches at school.
 - C. His grandfather was a football coach.
 - D. His grandfather works at a sports centre.

Part C: Short Answer

Answer using ONE WORD OR NUMBER.

9. How old is Ryan? _____
10. What instrument does Mia play? _____

Answer Key with Supporting Evidence

Item	Answer	Supporting Information	Reading Skill/Cognitive Process	Explanation
1	A	"My older cousin is teaching me some new computer skills."	Locate specific information	Jake receives help from a relative.
2	B	"We're practicing for a school	Locate specific information	Mia is preparing for a performance.

Item	Answer	Supporting Information	Reading Skill/Cognitive Process	Explanation
		concert next month.”		
3	A	“I’d like to work in software design.”	Identify future plans	Jake wants a technology-related career.
4	B	“I’m planning to attend a music camp.”	Locate specific information	Mia plans to attend a special course.
5	C	“I hope to join a regional football tournament.”	Locate specific information	Ryan hopes to participate in a sports competition.
6	C	“I’m taking part in a competition for young programmers.”	Locate details	Jake is improving his programming skills.
7	B	“We’re practicing for a school concert next month.”	Locate details	Mia is preparing for a concert.
8	C	“He used to coach a football team.”	Locate details	This explains why Ryan practises with him.
9	16	“I’m 16 years old.”	Scan for specific information	Learners locate a factual detail.
10	guitar	“I play the guitar.”	Scan for specific information	Learners identify Mia’s instrument.

Total Score: 10 marks.

Appendix G

B1+ Example of the Specification-Driven Approach: Test Specification, Assessment, and Text Inspector Results

Textbook: *Navigate Intermediate B1+* (Roberts, Buchanan, & Pathare, 2015)

Publisher: Oxford University Press

Target CEFR Level: B1+

Unit: Unit 1.1, *Are You Really My Friend?*

Source Reading Task: Learners read a continuous factual text describing Rob's attempt to meet hundreds of his Facebook friends in person and answer comprehension questions that require them to understand the main ideas, supporting details, and overall message of the text.

Selected CEFR Reading Descriptor: *Can understand straightforward factual texts on subjects related to their fields and interests with a satisfactory level of comprehension* (Council of Europe, 2020, p. 54).

Reading Test Specification

Specification Component	Recommended Specification	Rationale
Target CEFR Level	B1+ (Upper B1 / Intermediate)	The target descriptor requires learners to understand straightforward factual texts related to familiar interests with satisfactory comprehension. The chapter centers on a factual magazine-style article about friendships formed through social media and requires learners to identify explicit information, understand the author's purpose, and comprehend factual details.
Reading Construct	Ability to understand straightforward factual information in an authentic-style informational text by identifying main ideas, locating supporting details, understanding	The reading activities require learners to identify Rob Jones's motivations, recognize outcomes of his actions, understand factual information, interpret his views on friendship, and integrate information across the article.

Specification Component	Recommended Specification	Rationale
	reasons and purposes, recognizing factual relationships, and making simple text-based inferences.	These abilities align closely with the target CEFR descriptor.
Text Type	One informational magazine-style or online feature article describing a real person, social trend, or personal experience.	The chapter presents a single continuous factual article (“Face-to-face with Facebook friends”) describing Rob Jones’s experiences and opinions. A parallel assessment should therefore maintain the same authentic informational genre.
Topic Domain	Friendship, social relationships, social networking, communication, travel, community involvement, and everyday personal experiences.	The chapter focuses on friendship and changing social relationships through social media while also incorporating travel, charity work, and interpersonal communication. These familiar themes are appropriate for upper B1 learners.
Text Length	Approximately 280–350 words presented as one continuous informational text.	The reading material consists of a single magazine-style article of approximately 300 words. Maintaining a similar length and continuous structure preserves the reading demands and organizational characteristics of the original chapter.
Vocabulary and Grammar Expectations	Vocabulary should primarily consist of high-frequency B1+ vocabulary related to friendship, social media, travel, communication, charity, and personal experiences. Lexical diversity should be moderate with	The chapter explicitly integrates present simple, present continuous, and present perfect within the reading text and subsequent reading-focused activities. Vocabulary is familiar and topic-specific, with repeated lexical items (e.g., friends, Facebook, travel,

Specification Component	Recommended Specification	Rationale
	purposeful recycling of key vocabulary throughout the text. Grammar should include present simple, present continuous, present perfect, occasional present perfect continuous, and straightforward complex sentences using common cohesive devices. Grammar should support comprehension rather than increase linguistic difficulty.	charity, meet) supporting comprehension.
Item Format	Multiple-choice comprehension questions, short-answer questions, factual information questions, and one main-idea question assessing explicit understanding and simple interpretation.	The reading activities ask learners to identify factual information, understand reasons, interpret outcomes, and answer comprehension questions directly related to the article. These formats effectively assess reading comprehension while minimizing language production.
Number of Items	8–10 items	A set of 8–10 items provides sufficient evidence of learners' comprehension of the factual text while remaining proportional to the relatively short reading passage and comparable to the chapter activities.
Intended Cognitive Process	• Identifying explicitly stated information • Locating specific details • Understanding main ideas • Identifying reasons, purposes, and outcomes • Recognizing	The comprehension activities require learners to retrieve factual information, connect ideas across the article, understand motivations and outcomes, and synthesize information to answer both local and global comprehension

Specification Component	Recommended Specification	Rationale
	relationships between ideas • Making simple text-based inferences • Integrating information across paragraphs	questions. These processes align closely with the target CEFR descriptor.
Scoring Method	Dichotomous scoring (1 point for each correct response; 0 points for each incorrect response). Report both raw score and percentage correct.	The reading construct focuses on objective comprehension of factual information. Dichotomous scoring provides reliable evidence of successful comprehension and information retrieval.

Summary of Key Assessment Characteristics

- Target CEFR Level
 - B1+ (Upper B1 / Intermediate)
- Reading Construct
 - Ability to understand straightforward factual information in an authentic-style informational text by identifying main ideas, locating supporting details, understanding reasons and purposes, recognizing factual relationships, and making simple text-based inferences.
- Text Type
 - One informational magazine-style or online feature article describing a real person, social trend, or personal experience.
- Topic Domain
 - Friendship, social relationships, social networking, communication, travel, community involvement, and everyday personal experiences.
- Text Length
 - Approximately 280–350 words presented as one continuous informational text.
- Vocabulary and Grammar Expectations
 - High-frequency B1+ vocabulary related to friendship, social media, travel, communication, charity, and personal experiences.
 - Moderate lexical diversity with recycling of key topic vocabulary.

- Present simple, present continuous, present perfect, and occasional present perfect continuous.
 - Straightforward complex sentences and cohesive devices that support comprehension.
- Item Format
 - Multiple-choice questions, short-answer questions, factual information questions, and one main-idea question.
- Number of Items
 - 8–10 items.
- Intended Cognitive Process
 - Identifying explicitly stated information.
 - Locating specific details.
 - Understanding main ideas.
 - Identifying reasons, purposes, and outcomes.
 - Recognizing relationships between ideas.
 - Making simple text-based inferences.
 - Integrating information across paragraphs.
- Scoring Method
 - Dichotomous scoring (1 point per correct response; 0 points per incorrect response), reported as both a raw score and percentage correct.

Reading Passage

Making Connections in a New City

When Daniel Ruiz moved to a new city for work two years ago, he knew very few people. Although he enjoyed his job, he often spent his weekends alone. He wanted to meet new people, but he was not sure how to start.

One day, a colleague suggested joining a community website that advertises local activities and events. Daniel decided to try it. He signed up for a photography group, a cycling club, and a weekend volunteering team. Since then, he has attended more than forty events and has met many people from different backgrounds.

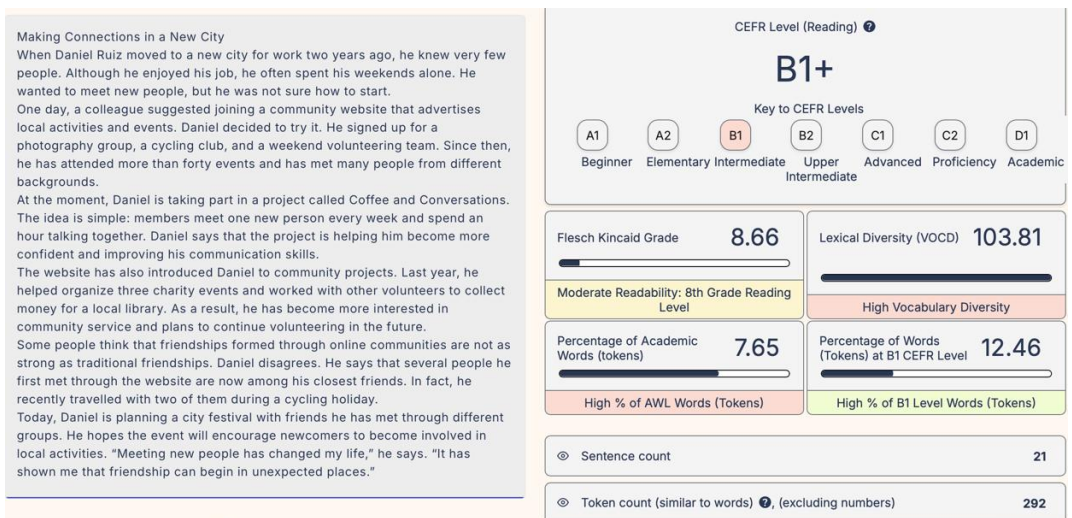
At the moment, Daniel is taking part in a project called *Coffee and Conversations*. The idea is simple: members meet one new person every week and spend an hour talking together. Daniel says that the project is helping him become more confident and improving his communication skills.

The website has also introduced Daniel to community projects. Last year, he helped organize three charity events and worked with other volunteers to

collect money for a local library. As a result, he has become more interested in community service and plans to continue volunteering in the future.

Some people think that friendships formed through online communities are not as strong as traditional friendships. Daniel disagrees. He says that several people he first met through the website are now among his closest friends. In fact, he recently travelled with two of them during a cycling holiday.

Today, Daniel is planning a city festival with friends he has met through different groups. He hopes the event will encourage newcomers to become involved in local activities. "Meeting new people has changed my life," he says. "It has shown me that friendship can begin in unexpected places."



Assessment Items

Multiple-Choice Questions

- Why did Daniel join the community website?
 - To find a better job
 - To learn photography professionally
 - To meet new people and become involved in activities
 - To organize a city festival
- Which group did Daniel join after signing up for the website?
 - A language-learning group
 - A photography group

- C. A music club
- D. A business network

3. What is Daniel currently doing?

- A. Organizing charity events
- B. Looking for a new apartment
- C. Taking part in the *Coffee and Conversations* project
- D. Working for the community website

4. What has Daniel gained from the Coffee and Conversations project?

- A. More free time
- B. Better communication skills and confidence
- C. A promotion at work
- D. New photography equipment

5. Why has Daniel become more interested in community service?

- A. He wants to change careers.
- B. His employer requires volunteer work.
- C. His experience helping with community projects was positive.
- D. He enjoys working alone.

6. What can be inferred about Daniel's opinion of friendships formed online?

- A. They are usually temporary.
- B. They are less meaningful than traditional friendships.
- C. They can develop into close and lasting friendships.
- D. They only exist on the internet.

Short-Answer Questions

7. Approximately how many events has Daniel attended since joining the website? _____

8. Name two activities or groups that Daniel joined through the website.

- 1) _____
- 2) _____

Main Idea Question

9. What is the main idea of the article?

- A. Community websites are replacing traditional friendships.
- B. Daniel used local activities and community projects to build friendships and become more involved in his new city.

C. Volunteering is the best way to improve communication skills.

D. Moving to a new city is always difficult.

Answer Key and Evidence

Item	Correct Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
1	C	Paragraph 1: “He wanted to meet new people...” Paragraph 2: “Daniel decided to try it.”	Identifying purpose	Daniel joined because he wanted to meet people and participate in activities.
2	B	Paragraph 2: “He signed up for a photography group...”	Locating specific details	The photography group is explicitly mentioned.
3	C	Paragraph 3: “At the moment, Daniel is taking part in a project called Coffee and Conversations.”	Identifying explicitly stated information	The answer is directly stated.
4	B	Paragraph 3: “helping him become more confident and improving his communication skills.”	Identifying outcomes	Students identify the results of participation in the project.
5	C	Paragraph 4: “As a result, he has become more interested in community service...”	Identifying cause-and-effect relationships	Students connect Daniel’s volunteer experience with his increased interest in community service.

Item	Correct Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
6	C	Paragraph 5: “several people he first met through the website are now among his closest friends.”	Making simple text-based inferences	The statement supports the inference that online friendships can become meaningful friendships.
7	More than forty events	Paragraph 2: “he has attended more than forty events.”	Locating specific details	The information is explicitly stated.
8	Any two: photography group, cycling club, weekend volunteering team	Paragraph 2	Locating specific details	Students identify activities mentioned in the text.
9	B	Information from Paragraphs 1–6	Understanding the main idea / Integrating information across the text	Students must synthesize information from the entire article rather than focus on a single detail.

Scoring: 1 point per correct answer. Total = 9 points.

Percentage score = $(\text{Total Correct} \div 9) \times 100$.

Appendix H

B1+ Example of the Baseline Approach:
Assessment and Text Inspector Results

Textbook: *Navigate Intermediate B1+* (Roberts, Buchanan, & Pathare, 2015)

Publisher: Oxford University Press

Target CEFR Level: B1+

Unit: Unit 1.1, *Are You Really My Friend?*

Source Reading Task: Learners read a continuous factual text describing

Rob's attempt to meet hundreds of his Facebook friends in person and answer comprehension questions that require them to understand the main ideas, supporting details, and overall message of the text.

Selected CEFR Reading Descriptor: *Can understand straightforward factual texts on subjects related to their fields and interests with a satisfactory level of comprehension* (Council of Europe, 2020, p. 54).

Reading Passage

Making Friends Through Shared Interests

Many people use social media to stay in touch with friends and family, but some people use it to create new friendships as well. One example is 27-year-old Maya Chen from Singapore.

Three years ago, Maya joined an online photography group. At first, she only shared pictures and commented on other members' work. However, after a few months, she started talking regularly with several people in the group. They exchanged ideas about photography, travel, and daily life.

Since then, Maya has met twelve members of the group in person. She has travelled to five countries, including Japan, Australia, and Germany, to attend photography events and meet her online friends. Although she was nervous before her first meeting, she says that all the people she met were friendly and welcoming.

Maya believes that shared interests help people form strong friendships. "When people enjoy the same activities, they already have something to talk about," she explains. She also thinks that online communication can help shy people become more confident.

Not everyone agrees with Maya's views. Some people argue that online friendships are less meaningful than face-to-face relationships because people spend less time together in real life. Others worry that it can be difficult to know whether people are honest online.

Despite these concerns, Maya continues to participate in the photography group. She says the experience has taught her about different cultures and helped her develop friendships with people she would never have met otherwise.

Making Friends Through Shared Interests

Many people use social media to stay in touch with friends and family, but some people use it to create new friendships as well. One example is 27-year-old Maya Chen from Singapore.

Three years ago, Maya joined an online photography group. At first, she only shared pictures and commented on other members' work. However, after a few months, she started talking regularly with several people in the group. They exchanged ideas about photography, travel, and daily life.

Since then, Maya has met twelve members of the group in person. She has travelled to five countries, including Japan, Australia, and Germany, to attend photography events and meet her online friends. Although she was nervous before her first meeting, she says that all the people she met were friendly and welcoming.

Maya believes that shared interests help people form strong friendships. "When people enjoy the same activities, they already have something to talk about," she explains. She also thinks that online communication can help shy people become more confident.

Not everyone agrees with Maya's views. Some people argue that online friendships are less meaningful than face-to-face relationships because people spend less time together in real life. Others worry that it can be difficult to know whether people are honest online.

Despite these concerns, Maya continues to participate in the photography group. She says the experience has taught her about different cultures and helped her develop friendships with people she would never have met otherwise.

CEFR Level (Reading)

B1+

Key to CEFR Levels

A1	A2	B1	B2	C1	C2	D1
Beginner	Elementary	Intermediate	Upper Intermediate	Advanced	Proficiency	Academic

Flesch Kincaid Grade 8.97	Lexical Diversity (VOC) 115.46
Moderate Readability: 8th Grade Reading Level	High Vocabulary Diversity
Percentage of Academic Words (tokens) 4.49	Percentage of Words (Tokens) at B1 CEFR Level 13.01
High % of AWL Words (Tokens)	High % of B1 Level Words (Tokens)

🔍 Sentence count **17**

🔍 Token count (similar to words) , (excluding numbers) **251**

Copy Text

Questions

- Why did Maya join the online photography group?
 - To find a new job
 - To share and discuss photography
 - To learn a foreign language
 - To sell her photographs
- How many group members has Maya met in person?
 - Five
 - Ten
 - Twelve
 - Twenty
- What was Maya's feeling before meeting her online friends for the first time?
 - Excited
 - Bored
 - Nervous
 - Angry
- According to Maya, why can shared interests help people become friends?
 - They make people travel more often.
 - They provide common topics for conversation.
 - They help people earn more money.
 - They encourage people to spend less time online.

5. What benefit of online communication does Maya mention?
 - A. It helps shy people become more confident.
 - B. It improves people's photography skills immediately.
 - C. It allows people to avoid face-to-face meetings.
 - D. It makes people more popular.

6. What concern do some people have about online friendships?
 - A. They are too expensive to maintain.
 - B. They require travelling to other countries.
 - C. People may not always be honest online.
 - D. They prevent people from developing hobbies.

7. What has Maya learned from her experience?
 - A. How to become a professional photographer.
 - B. How to start a business abroad.
 - C. About different cultures and people.
 - D. About university life in other countries.

8. Which statement best summarizes the passage?
 - A. Social media should replace face-to-face communication.
 - B. Online friendships can create meaningful opportunities, although some people have concerns about them.
 - C. Photography is the best hobby for making friends.
 - D. Travelling is necessary for building friendships.

Answer Key with Supporting Evidence

Q	Correct Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
1	B	Paragraph 2: "At first, she only shared pictures and commented on other members' work."	Identifying explicit information	The text directly states Maya joined the group to share photography-related content.
2	C	Paragraph 3: "Since then, Maya has met twelve members of the group in person."	Locating specific details	The number is stated explicitly.

Q	Correct Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
3	C	Paragraph 3: "Although she was nervous before her first meeting..."	Understanding specific information	The text directly describes her feelings.
4	B	Paragraph 4: "They already have something to talk about."	Understanding stated reasons	Maya explains that shared interests provide common conversation topics.
5	A	Paragraph 4: "Online communication can help shy people become more confident."	Identifying explicit information	The benefit is directly mentioned.
6	C	Paragraph 5: "Others worry that it can be difficult to know whether people are honest online."	Understanding viewpoints	Learners identify a concern expressed by others.
7	C	Paragraph 6: "The experience has taught her about different cultures..."	Identifying main supporting details	The passage explicitly states what she learned.
8	B	Paragraphs 4–6 (overall message)	Understanding the main idea	The text presents both benefits and concerns about online friendships, making option B the best summary.

Appendix I

B2 Example of the Specification-Driven Approach: Test Specification, Assessment, and Text Inspector Results

Textbook: *Prepare Level 7* (Styring, Tims, & Chilton, 2019)

Publisher: Cambridge University Press

Target CEFR Level: B2

Unit: Unit 1, *Creative Minds*

Source Reading Task: Learners read four short biographies and complete a multiple-matching task by locating relevant information, distinguishing among similar experiences and achievements, and comparing information across multiple texts.

Selected CEFR Reading Descriptor: *Can scan quickly through long and complex texts, locating relevant details* (Council of Europe, 2020, p. 55).

Reading Test Specification

Specification Component	Recommended Specification	Rationale
Target CEFR Level	B2	The target descriptor requires learners to scan quickly through longer and more complex texts to locate relevant details. The chapter presents four biographical profiles of talented young people and requires learners to identify which person matches a series of statements by scanning and comparing information across texts. This directly reflects B2 reading demands.
Reading Construct	Ability to scan multiple texts efficiently to locate specific information, identify relevant details, compare information across texts, recognize explicit information	The reading activities require learners to search across four biographical texts to determine which individual matches each description. Successful performance depends on rapid information retrieval

Specification Component	Recommended Specification	Rationale
	about experiences, achievements, motivations, and future intentions, and integrate information from different texts to complete a task.	and comparison across texts rather than detailed linear reading.
Text Type	A set of four short biographical or informational profiles describing individuals, their achievements, experiences, interests, and future plans.	The reading passage consists of four independent profiles of successful young people from different creative fields. Maintaining a multiple-text format preserves the reading construct and reflects the chapter's organization.
Topic Domain	Creativity, innovation, arts, media, technology, entrepreneurship, personal achievement, careers, talents, and future aspirations.	The chapter theme is <i>Creative Minds</i> . The reading texts focus on young people who have achieved success in literature, music, online media, and technological innovation. These interconnected themes define the instructional focus of the unit.
Text Length	Approximately 500–700 words distributed across four short profiles, with each profile containing approximately 120–180 words.	The chapter presents four separate texts that together create a moderate reading load. Preserving both the approximate total length and the multiple-text structure maintains the intended scanning demands rather than relying on generic B2 expectations.
Vocabulary and Grammar Expectations	Vocabulary should consist primarily of upper-intermediate (B2) vocabulary related to	The chapter contains authentic upper-intermediate vocabulary (e.g., <i>genre</i> , <i>proposals</i> ,

Specification Component	Recommended Specification	Rationale
	creativity, media, publishing, music, technology, innovation, entrepreneurship, and achievement. Lexical diversity should be moderate to moderately high, with key topic vocabulary recycled across the four profiles. Grammar should include present simple, present continuous, present perfect, past simple, past perfect, relative clauses, complex noun phrases, and cause-and-effect relationships. These language features should support comprehension while maintaining an appropriate level of challenge for B2 learners.	<i>convinced, passion</i>) and varied grammatical structures describing achievements, experiences, and future plans. These features contribute to reading comprehension while remaining consistent with B2 expectations.
Item Format	Multiple-matching items requiring learners to match statements, descriptions, experiences, opinions, achievements, motivations, and future plans to one of several texts.	The chapter employs a multiple-matching format comparable to Cambridge B2 First Reading and Use of English Part 7. This format directly measures the ability to scan efficiently across multiple texts.
Number of Items	10–12 multiple-matching items	The chapter provides ten matching questions linked to four texts. A parallel assessment should contain a similar number of items to provide sufficient evidence of the intended reading construct while maintaining comparable

Specification Component	Recommended Specification	Rationale
		cognitive demands and testing time.
Intended Cognitive Process	• Scanning for specific information • Locating relevant details rapidly • Identifying explicit factual information • Recognizing references to experiences, achievements, motivations, and future intentions • Comparing information across multiple texts • Distinguishing similar but non-identical details	Learners must retrieve targeted information rapidly, compare information across four profiles, distinguish between similar descriptions, and identify evidence relating to achievements, interests, and future goals. These processes correspond closely to the target CEFR descriptor and the chapter's reading activities.
Scoring Method	Dichotomous scoring (1 point for each correct response; 0 points for each incorrect response). Report both the total score and percentage correct.	Each matching item has one objectively correct response. Dichotomous scoring provides reliable evidence of learners' ability to locate and compare information efficiently across multiple texts.

Summary of Key Assessment Characteristics

- Target CEFR Level
 - B2
- Reading Construct
 - Ability to scan multiple texts efficiently to locate specific information, identify relevant details, compare information across texts, recognize explicit information about experiences, achievements, motivations, and future intentions, and integrate information from different texts to complete a task.
- Text Type
 - Four short biographical or informational profiles describing individuals, their achievements, experiences, interests, and future plans.

- Topic Domain
 - Creativity, innovation, arts, media, technology, entrepreneurship, personal achievement, careers, talents, and future aspirations.
- Text Length
 - Approximately 500–700 words across four short profiles (approximately 120–180 words per profile).
- Vocabulary and Grammar Expectations
 - Upper-intermediate (B2) vocabulary related to creativity, media, publishing, music, technology, innovation, entrepreneurship, and achievement.
 - Moderate to moderately high lexical diversity with recycling of key topic vocabulary.
 - Present simple, present continuous, present perfect, past simple, past perfect, relative clauses, complex noun phrases, and cause-and-effect relationships that support comprehension.
- Item Format
 - Multiple-matching items requiring learners to match statements to one of several texts.
- Number of Items
 - 10–12 items.
- Intended Cognitive Process
 - Scanning for specific information.
 - Locating relevant details rapidly.
 - Identifying explicit factual information.
 - Recognizing experiences, achievements, motivations, and future intentions.
 - Comparing information across multiple texts.
 - Distinguishing similar but non-identical details.
- Scoring Method
 - Dichotomous scoring (1 point per correct response; 0 points per incorrect response), reported as both a raw score and percentage.

Reading Passage

INNOVATORS OF TOMORROW

A. Maya Rodriguez

Maya Rodriguez is a young environmental entrepreneur from Mexico. When she was fourteen, she became concerned about the large amount of food waste produced by local restaurants. After researching the problem,

she designed a mobile application that allows restaurants to sell surplus food at reduced prices instead of throwing it away. The project began as a school assignment, but Maya continued developing it after receiving positive feedback from local business owners. Within two years, more than 150 restaurants had joined the platform.

Maya has already won several national innovation awards, and she is currently working with universities to improve the technology behind the app. Although she enjoys studying engineering, she hopes to establish her own international company in the future and expand the service to other countries.

B. Leo Fischer

Leo Fischer, a photographer from Germany, first became interested in wildlife after joining a nature club at primary school. At the age of sixteen, he borrowed a camera from a friend and started documenting endangered animals in local forests. His photographs soon attracted attention on social media because they showed unusual animal behaviour that few people had seen before.

As his audience grew, Leo decided to organize free photography workshops for teenagers who wanted to learn more about nature conservation. He believes that photography can encourage people to protect the environment. Last year, one of his photo collections was displayed in a major museum, and he plans to publish a book featuring his work next year.

C. Priya Nair

Priya Nair is a software developer from India who discovered her passion for technology at an early age. By the time she was fifteen, she had already created several educational games designed to help children improve their mathematics skills. While attending secondary school, she noticed that many students struggled to find affordable learning resources.

As a result, Priya developed an online platform that provides free educational materials and interactive lessons. The website became widely used during periods of remote learning and now supports students in several countries. Priya regularly speaks at educational conferences and is currently leading a team that is developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality around the world.

D. Ethan Brooks

Ethan Brooks is a British musician and content creator. He had always enjoyed writing songs, but his career changed when he began posting short music tutorials online. Initially, he uploaded videos simply to share ideas with other musicians. However, his practical advice and friendly presentation style quickly attracted millions of viewers.

Ethan later launched a series of online courses for beginner musicians. Many participants reported that the courses helped them gain confidence and improve their skills. In addition to creating educational content, Ethan performs regularly at music festivals and charity events. He is currently preparing a global tour and plans to establish a foundation that will provide music education for young people from disadvantaged backgrounds.

Import Text
CEFR Level (Reading): Calculated/Target C2/B2

INNOVATORS OF TOMORROW

A. Maya Rodriguez

Maya Rodriguez is a young environmental entrepreneur from Mexico. When she was fourteen, she became concerned about the large amount of food waste produced by local restaurants. After researching the problem, she designed a mobile application that allows restaurants to sell surplus food at reduced prices instead of throwing it away. The project began as a school assignment, but Maya continued developing it after receiving positive feedback from local business owners. Within two years, more than 150 restaurants had joined the platform.

Maya has already won several national innovation awards, and she is currently working with universities to improve the technology behind the app. Although she enjoys studying engineering, she hopes to establish her own international company in the future and expand the service to other countries.

B. Leo Fischer

Leo Fischer, a photographer from Germany, first became interested in wildlife after joining a nature club at primary school. At the age of sixteen, he borrowed a camera from a friend and started documenting endangered animals in local forests. His photographs soon attracted attention on social media because they showed unusual animal behaviour that few people had seen before.

As his audience grew, Leo decided to organise free photography workshops for teenagers who wanted to learn more about nature conservation. He believes that photography can encourage people to protect the environment. Last year, one of his photo collections was displayed in a major museum, and he plans to publish a book featuring his work next year.

Click on one of the measures in the boxes below for more details.

CEFR Level (Reading)

C2

Key to CEFR Levels

A1	A2	B1	B2	C1	C2	D1	D2
Beginner	Elementary	Intermediate	Upper Intermediate	Advanced	Proficiency	Academic	Upper Academic

Flesch Kincaid Grade

12.98

Moderate Readability: 12th Grade Reading Level

Lexical Diversity (VOC)

172.67

High Vocabulary Diversity

Percentage of Academic Words (tokens)

11.68

High % of AWL Words (Tokens)

Percentage of Words (Tokens) at B1 CEFR Level

14.97

High % of B1 Level Words (Tokens)

Sentence count

28

Token count (similar to words) (excluding numbers)

490

23C179B8-8EA5-11F1-B30E-E778C44C6...

Import Text

CEFR Level (Reading): Calculated/Target C2/B2

INNOVATORS OF TOMORROW

A. Maya Rodriguez

Maya Rodriguez is a young environmental entrepreneur from Mexico. When she was fourteen, she became concerned about the large amount of food waste produced by local restaurants. After researching the problem, she designed a mobile application that allows restaurants to sell surplus food at reduced prices instead of throwing it away. The project began as a school assignment, but Maya continued developing it after receiving positive feedback from local business owners. Within two years, more than 150 restaurants had joined the platform.

Maya has already won several national innovation awards, and she is currently working with universities to improve the technology behind the app. Although she enjoys studying engineering, she hopes to establish her own international company in the future and expand the service to other countries.

B. Leo Fischer

Leo Fischer, a photographer from Germany, first became interested in wildlife after joining a nature club at primary school. At the age of sixteen, he borrowed a camera from a friend and started documenting endangered animals in local forests. His photographs soon attracted attention on social media because they showed unusual animal behaviour that few people had seen before.

As his audience grew, Leo decided to organise free photography workshops for teenagers who wanted to learn more about nature conservation. He believes that photography can encourage people to protect the environment. Last year, one of his photo collections was displayed in a major museum, and he plans to publish a book featuring his work next year.

C. Priya Nair

Priya Nair is a software developer from India who discovered her passion for technology at an early age. By the time she was fifteen, she had already created several educational games designed to help children improve their mathematics skills. While attending secondary school, she noticed that many students struggled to find affordable learning resources.

As a result, Priya developed an online platform that provides free educational materials and interactive lessons. The website became widely used during periods of remote learning and now supports students in several countries. Priya regularly speaks at educational conferences and is currently leading a team that is developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality around the world.

her own international company in the future and expand the service to other countries.

B. Leo Fischer

Leo Fischer, a photographer from Germany, first became interested in wildlife after joining a nature club at primary school. At the age of sixteen, he borrowed a camera from a friend and started documenting endangered animals in local forests. His photographs soon attracted attention on social media because they showed unusual animal behaviour that few people had seen before.

As his audience grew, Leo decided to organise free photography workshops for teenagers who wanted to learn more about nature conservation. He believes that photography can encourage people to protect the environment. Last year, one of his photo collections was displayed in a major museum, and he plans to publish a book featuring his work next year.

C. Priya Nair

Priya Nair is a software developer from India who discovered her passion for technology at an early age. By the time she was fifteen, she had already created several educational games designed to help children improve their mathematics skills. While attending secondary school, she noticed that many students struggled to find affordable learning resources.

As a result, Priya developed an online platform that provides free educational materials and interactive lessons. The website became widely used during periods of remote learning and now supports students in several countries. Priya regularly speaks at educational conferences and is currently leading a team that is developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality around the world.

D. Ethan Brooks

Ethan Brooks is a British musician and content creator. He had always enjoyed writing songs, but his career changed when he began posting short music tutorials online. Initially, he uploaded videos simply to share ideas with other musicians. However, his practical advice and friendly presentation style quickly attracted millions of viewers.

Ethan later launched a series of online courses for beginner musicians. Many participants reported that the courses helped them gain confidence and improve their skills. In addition to creating educational content, Ethan performs regularly at music festivals and charity events. He is currently preparing a global tour and plans to establish a foundation that will provide music education for young people from disadvantaged backgrounds.

Lexis: English Vocabulary Profile
(American English)

Analysis Options

This tool assigns each word a CEFR level, according to the English Vocabulary Profile from Cambridge University Press. You can change the English type in the Analysis Options above.

If you want to correct the analysis, click on the "✎" icon next to a word below, change the tag, then click Update.

Word List

Word Charts

Lexis: EVP Export

See All	CEFR	Unique Words	Words	Cumul % Unique Words	Cumul % Words
See All	A1	104	249	35.74%	51.77%
See All	A2	57	72	19.59%	14.97%
See All	B1	65	72	22.34%	14.97%
See All	B2	29	37	9.97%	7.69%
See All	C1	8	8	2.75%	1.66%
See All	C2	3	3	1.03%	0.62%
See All	Unlisted	25	40	8.59%	8.32%

Lexis: English Vocabulary Profile
(American English)

Analysis Options

This tool assigns each word a CEFR level, according to the English Vocabulary Profile from Cambridge University Press. You can change the English type in the Analysis Options above.

If you want to correct the analysis, click on the "✎" icon next to a word below, change the tag, then click Update.

Word List

Word Charts

Lexis: EVP Export

See All	CEFR	Unique Words	Words	Cumul % Unique Words	Cumul % Words
See All	A1	104	249	35.74%	51.77%
See All	A2	57	72	19.59%	14.97%
See All	B1	65	72	22.34%	14.97%
See All	B2	29	37	9.97%	7.69%
See All	C1	8	8	2.75%	1.66%
See All	C2	3	3	1.03%	0.62%
See All	Unlisted	25	40	8.59%	8.32%

Assessment Items
Questions 1–12

Read the texts about four innovators. Which person (A–D) matches each statement?

The people may be chosen more than once.

1. This person developed a project after identifying a community problem.
2. This person first became interested in their field through participation in an organized activity.
3. This person created something to improve access to education.
4. This person intends to start a larger organization in the future.
5. This person uses their success to teach or support younger people.
6. This person's work gained public attention through online platforms.
7. This person hopes to address a global social issue through their work.
8. This person plans to publish a major piece of work in the future.
9. This person transformed an activity originally done for personal reasons into a successful career.
10. This person continued a project that originally began in an educational setting.
11. This person is currently collaborating with other institutions.
12. This person believes their work can encourage positive behaviour in society.

Answer Key and Evidence

Item	Answer	Supporting Evidence	Reading Skill / Cognitive Process	Explanation
1	A	"became concerned about the large amount of food waste" and "designed a mobile application"	Scanning for specific information; identifying explicit factual information	Maya identified a local problem and created a solution.
2	B	"first became interested in wildlife after joining a nature club"	Locating relevant details	Leo's interest developed through participation in an organized club.
3	C	"developed an online platform that provides	Identifying explicit factual information	Priya created a resource that improves

Item	Answer	Supporting Evidence	Reading Skill / Cognitive Process	Explanation
		free educational materials”		educational access.
4	A	“hopes to establish her own international company in the future”	Recognizing future plans	Maya intends to create a larger organization.
5	B	“organize free photography workshops for teenagers”	Locating relevant details; recognizing motivations	Leo supports younger people through workshops.
6	B	“photographs soon attracted attention on social media”	Scanning for specific information	Leo gained recognition through online platforms.
7	C	“hopes her work will reduce educational inequality around the world”	Recognizing future intentions and goals	Priya aims to address a global social issue.
8	B	“plans to publish a book featuring his work next year”	Recognizing future plans	Leo intends to publish a book.
9	D	“uploaded videos simply to share ideas” and later attracted millions of viewers	Comparing information; identifying achievements	Ethan turned a personal hobby into a successful career.
10	A	“The project began as a school assignment, but	Locating relevant details	Maya’s project originated in an educational context.

Item	Answer	Supporting Evidence	Reading Skill / Cognitive Process	Explanation
		Maya continued developing it”		
11	A	“currently working with universities to improve the technology”	Scanning for specific information	Maya is collaborating with institutions.
12	B	“photography can encourage people to protect the environment”	Identifying explicit opinions and motivations	Leo believes his work can influence positive behaviour.

Appendix J

B2 Example of the Baseline Approach: Assessment and Text Inspector Results

Textbook: *Prepare Level 7* (Styring, Tims, & Chilton, 2019)

Publisher: Cambridge University Press

Target CEFR Level: B2

Unit: Unit 1, *Creative Minds*

Source Reading Task: Learners read four short biographies and complete a multiple-matching task by locating relevant information, distinguishing among similar experiences and achievements, and comparing information across multiple texts.

Selected CEFR Reading Descriptor: *Can scan quickly through long and complex texts, locating relevant details* (Council of Europe, 2020, p. 55).

Reading Passage

YOUNG CHANGEMAKERS

A. Elena Petrova

When Elena Petrova was thirteen, she became interested in environmental issues after watching a documentary about plastic pollution in the oceans. She began researching alternatives to single-use plastics and eventually

designed a reusable food container made from plant-based materials. At first, her parents thought the project would remain a school activity and doubted whether it could be developed further. However, Elena continued refining her design and entered several science competitions. Her invention received national media attention after she won a major youth innovation award. Elena is now working with manufacturers to produce the containers commercially and hopes they will become widely available within the next few years.

B. Marcus Reed

Marcus Reed developed a passion for photography while travelling with his family during summer holidays. Wanting to improve his skills, he enrolled in an online photography course and spent hours practicing techniques used by professional photographers. He began sharing his photographs on social media and quickly attracted a large audience. One of his nature photographs was featured in an international magazine, bringing him wider recognition. Today, Marcus runs workshops for teenagers who are interested in photography and often says that helping others develop creative skills is the most rewarding part of his work. Although he enjoys his current projects, he has not yet decided exactly what direction his career will take.

C. Hana Kim

Hana Kim started learning computer programming when she was only nine years old. She mastered several coding languages within a few years and soon began creating educational games for children. Her interest in educational technology developed after she noticed how difficult some younger students found mathematics lessons. To better understand classroom challenges, she volunteered as a teaching assistant at a local learning centre. One of her educational apps became extremely popular and was downloaded more than five million times in a single year, setting a record among student-developed applications in her country. Hana plans to continue developing technology that supports learning.

D. Lucas Moreno

Lucas Moreno always enjoyed music, but his interest in sound engineering began after attending a concert and observing the technical team behind the performance. Determined to learn more, he took apart old audio equipment and taught himself how recording systems worked. By the age of sixteen, he had already designed a low-cost recording device for young musicians. His teachers immediately recognized his talent, although Lucas himself was uncertain whether his invention would succeed. The device later received several industry awards. Lucas frequently visits schools to encourage

students to explore careers in science, technology, and the arts, and he hopes to establish a training centre for young innovators in the future.

1DB5CD78-6EB2-11F1-B30E-E778C...

Summary Analysis Options

Click on one of the measures in the boxes below for more details.

CEFR Level (Reading) ?

C1+

Key to CEFR Levels

A1 Beginner A2 Elementary B1 Intermediate B2 Upper Intermediate C1 Advanced C2 Proficiency D1 Academic D2 Upper Academic

Flesch Kincaid Grade **13.56**

Lexical Diversity (VOC) **185.23**

Low Readability: University Student Reading Level

High Vocabulary Diversity

Percentage of Academic Words (tokens) **11.19**

Percentage of Words (Tokens) at B1 CEFR Level **15.42**

High % of AWL Words (Tokens)

High % of B1 Level Words (Tokens)

Ⓢ Sentence count **24**

Ⓢ Token count (similar to words) ? (excluding numbers) **447**

YOUNG CHANGEMAKERS

A. Elena Petrova

When Elena Petrova was thirteen, she became interested in environmental issues after watching a documentary about plastic pollution in the oceans. She began researching alternatives to single-use plastics and eventually designed a reusable food container made from plant-based materials. At first, her parents thought the project would remain a school activity and doubted whether it could be developed further. However, Elena continued refining her design and entered several science competitions. Her invention received national media attention after she won a major youth innovation award. Elena is now working with manufacturers to produce the containers commercially and hopes they will become widely available within the next few years.

B. Marcus Reed

Marcus Reed developed a passion for photography while travelling with his family during summer holidays. Wanting to improve his skills, he enrolled in an online photography course and spent hours practising techniques used by professional photographers. He began sharing his photographs on social media and quickly attracted a large audience. One of his nature photographs was featured in an international magazine, bringing him wider recognition. Today, Marcus runs workshops for teenagers who are interested in photography and often says that helping others develop creative skills is the most rewarding part of his work. Although he enjoys his current projects, he has not yet decided exactly what direction his career will take.

C. Hana Kim

Hana Kim started learning computer programming when she was only nine years old. She mastered several coding languages within a few years and soon began creating educational games for children. Her interest in educational technology developed after she noticed how difficult some younger students found mathematics lessons. To better understand classroom challenges, she volunteered as a teaching assistant at a local learning centre. One of her educational apps became extremely popular and was downloaded more than five million times in a single year, setting a record among student-developed applications in her country. Hana plans to continue developing technology that supports learning.

D. Lucas Moreno

Lucas Moreno always enjoyed music, but his interest in sound engineering began after attending a concert and observing the technical team behind the performance. Determined to learn more, he took apart old audio equipment and taught himself how recording systems worked. By the age of sixteen, he had already designed a low-cost recording device for young musicians. His teachers immediately recognised his talent, although Lucas himself was uncertain whether his invention would succeed. The device later received several industry awards. Lucas frequently visits schools to encourage students to explore careers in science, technology, and the arts, and he hopes to establish a training centre for young innovators in the future.

YOUNG CHANGEMAKERS

A. Elena Petrova

When Elena Petrova was thirteen, she became interested in environmental issues after watching a documentary about plastic pollution in the oceans. She began researching alternatives to single-use plastics and eventually designed a reusable food container made from plant-based materials. At first, her parents thought the project would remain a school activity and doubted whether it could be developed further. However, Elena continued refining her design and entered several science competitions. Her invention received national media attention after she won a major youth innovation award. Elena is now working with manufacturers to produce the containers commercially and hopes they will become widely available within the next few years.

B. Marcus Reed

Marcus Reed developed a passion for photography while travelling with his family during summer holidays. Wanting to improve his skills, he enrolled in an online photography course and spent hours practising techniques used by professional photographers. He began sharing his photographs on social media and quickly attracted a large audience. One of his nature photographs was featured in an international magazine, bringing him wider recognition. Today, Marcus runs bookshops for teenagers who are interested in photography and often says that helping others develop creative skills is the most rewarding part of his work. Although he enjoys his current projects, he has not yet decided exactly what direction his career will take.

C. Hana Kim

Hana Kim started learning computer programming when she was only nine years old. She mastered several coding languages within a few years and soon began creating educational games for children. Her interest in educational technology developed after she noticed how difficult some younger students found mathematics lessons. To better understand classroom challenges, she volunteered as a teaching assistant at a local learning centre. One of her educational apps became extremely popular and was downloaded more than five million times in a single year, setting a record among student-developed applications in her country. Hana plans to continue developing technology that supports learning.

D. Lucas Moreno

Lucas Moreno always enjoyed music, but his interest in sound engineering began after attending a concert and observing the technical team behind the performance. Determined to learn more, he took apart old audio equipment and taught himself how recording systems worked. By the age of sixteen, he had already designed a low-cost recording device for young musicians. His teachers immediately recognised his talent, although Lucas himself was uncertain whether his invention would succeed. The device later received several industry awards. Lucas frequently visits schools to encourage students to explore careers in science, technology, and the arts, and he hopes to establish a training centre for young innovators in the future.

Lexis: English Vocabulary Profile (American English) Analysis Options

This tool assigns each word a CEFR level, according to the English Vocabulary Profile from Cambridge University Press. You can change the English type in the Analysis Options above.

If you want to correct the analysis, click on the "Ⓢ" icon next to a word below, change the tag, then click Update.

Word List Word Charts

Lexis: EVP Export

See All	CEFR	Unique Words	Words	Cumul % Unique Words	Cumul % Words
Ⓢ	A1	111	224	39.22%	50.79%
Ⓢ	A2	63	78	22.26%	17.69%
Ⓢ	B1	56	68	19.79%	15.42%
Ⓢ	B2	25	29	8.83%	6.58%
Ⓢ	C1	3	3	1.06%	0.68%
Ⓢ	C2	2	2	0.71%	0.45%
Ⓢ	Unlisted	23	37	8.13%	8.39%

Show All

Questions

For questions 1–10, choose the person (A–D). The people may be chosen more than once.

Which person ...

1. developed an interest after observing other people at work?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

2. wanted practical experience in order to understand a situation better?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

3. achieved a record with their activity?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

4. had family members who were initially uncertain about their project?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

5. is involved in helping other people develop their abilities?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

6. has not yet made a firm decision about future career plans?
 - A. Elena
 - B. Marcus
 - C. Hana
 - D. Lucas

7. gained recognition through the media?
 - A. Elena
 - B. Marcus

- C. Hana
D. Lucas

8. learned a complex skill at a young age?
A. Elena
B. Marcus
C. Hana
D. Lucas
9. has a clear idea about what they want to do in the future?
A. Elena
B. Marcus
C. Hana
D. Lucas
10. undertook additional learning to improve their performance?
A. Elena
B. Marcus
C. Hana
D. Lucas

Answer Key with Supporting Evidence

Q	Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
1	D	“interest in sound engineering began after attending a concert and observing the technical team behind the performance”	Scanning for specific detail	Lucas became interested after watching others work.
2	C	“To better understand classroom challenges, she volunteered as a teaching assistant”	Scanning for purpose	Hana sought firsthand experience to understand a problem.

Q	Answer	Supporting Information	Reading Skill / Cognitive Process	Explanation
3	C	“downloaded more than five million times... setting a record”	Scanning for factual detail	Hana’s app achieved a record.
4	A	“her parents... doubted whether it could be developed further”	Identifying specific information	Elena’s family initially doubted her project.
5	B	“Marcus runs workshops for teenagers”	Scanning for relevant detail	He helps others develop photography skills.
6	B	“he has not yet decided exactly what direction his career will take”	Identifying future intentions	Marcus remains undecided about his future career.
7	A	“received national media attention”	Scanning for recognition/ publicity	Elena was recognized through media coverage.
8	C	“started learning computer programming when she was only nine years old”	Locating biographical detail	Hana acquired an advanced skill at a young age.
9	D	“he hopes to establish a training centre for young innovators in the future”	Identifying future plans	Lucas has a clearly stated future goal.
10	B	“enrolled in an online photography course and spent hours practicing”	Scanning for actions taken to improve	Marcus undertook additional training to improve his skills.

Appendix K

Revised B2 Reading Assessment Generated Using Prompt C2 and Corresponding Text Inspector Results

INNOVATORS OF TOMORROW

A. Maya Rodriguez

Maya Rodriguez is a young entrepreneur from Mexico. At fourteen, she became concerned about food waste in local restaurants. She created a mobile app that allows restaurants to sell extra food at lower prices instead of throwing it away. The project started as a school assignment, but Maya continued developing it. Today, more than 150 restaurants use the app.

Maya has won several innovation awards and is currently working with universities to improve the technology. She hopes to create an international company and expand the service to other countries.

B. Leo Fischer

Leo Fischer is a wildlife photographer from Germany. He became interested in nature after joining a school nature club. At sixteen, he began photographing endangered animals. His work gained attention on social media because it showed unusual animal behaviour.

Leo now runs free photography workshops for teenagers. He believes photography can encourage people to protect the environment. He plans to publish a book of his photographs next year.

C. Priya Nair

Priya Nair is a software developer from India. As a teenager, she created educational games to help children improve their mathematics skills. Later, she developed a website that provides free learning materials and online lessons.

The platform is now used by students in several countries. Priya currently leads a team developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality worldwide.

D. Ethan Brooks

Ethan Brooks is a British musician and content creator. He started posting music tutorials online to share ideas with other musicians. His videos soon attracted millions of viewers.

Ethan later created online courses for beginner musicians. He performs at festivals and charity events and is preparing for a global tour. He also plans to establish a foundation that will provide music education for disadvantaged young people.

INNOVATORS OF TOMORROW

A. Maya Rodriguez
Maya Rodriguez is a young entrepreneur from Mexico. At fourteen, she became concerned about food waste in local restaurants. She created a mobile app that allows restaurants to sell extra food at lower prices instead of throwing it away. The project started as a school assignment, but Maya continued developing it. Today, more than 150 restaurants use the app. Maya has won several innovation awards and is currently working with universities to improve the technology. She hopes to create an international company and expand the service to other countries.

B. Leo Fischer
Leo Fischer is a wildlife photographer from Germany. He became interested in nature after joining a school nature club. At sixteen, he began photographing endangered animals. His work gained attention on social media because it showed unusual animal behaviour. Leo now runs free photography workshops for teenagers. He believes photography can encourage people to protect the environment. He plans to publish a book of his photographs next year.

C. Priya Nair
Priya Nair is a software developer from India. As a teenager, she created educational games to help children improve their mathematics skills. Later, she developed a website that provides free learning materials and online lessons. The platform is now used by students in several countries. Priya currently leads a team developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality worldwide.

D. Ethan Brooks
Ethan Brooks is a British musician and content creator. He started posting music tutorials online to share ideas with other musicians. His videos soon attracted millions of viewers. Ethan later created online courses for beginner musicians. He performs at festivals and charity events and is preparing for a global tour. He also plans to establish a foundation that will provide music education for disadvantaged young people.

CEFR Level (Reading) **C1**

Key to CEFR Levels

A1 Beginner A2 Elementary B1 Intermediate B2 Upper Intermediate C1 Advanced C2 Proficiency D1 Academic D2 Upper Academic

Flesch Kincaid Grade **10.55**

Moderate Readability: 10th Grade Reading Level

Lexical Diversity (VOCd) **151.41**

High Vocabulary Diversity

Percentage of Academic Words (tokens) **9.04**

High % of AWL Words (Tokens)

Percentage of Words (Tokens) at B1 CEFR Level **12.46**

High % of B1 Level Words (Tokens)

Sentence count **26**

Token count (similar to words) **302**

Possible issues: 4

Our research suggests that these statistics are higher than would typically be seen in a B2 Level Reading text.

FOB18998-6EA7-11F1-B30E-E778C...

Import Text

CEFR Level (Reading): Calculated/Target C1/B2

INNOVATORS OF TOMORROW

A. Maya Rodriguez
Maya Rodriguez is a young entrepreneur from Mexico. At fourteen, she became concerned about food waste in local restaurants. She created a mobile app that allows restaurants to sell extra food at lower prices instead of throwing it away. The project started as a school assignment, but Maya continued developing it. Today, more than 150 restaurants use the app. Maya has won several innovation awards and is currently working with universities to improve the technology. She hopes to create an international company and expand the service to other countries.

B. Leo Fischer
Leo Fischer is a wildlife photographer from Germany. He became interested in nature after joining a school nature club. At sixteen, he began photographing endangered animals. His work gained attention on social media because it showed unusual animal behaviour. Leo now runs free photography workshops for teenagers. He believes photography can encourage people to protect the environment. He plans to publish a book of his photographs next year.

C. Priya Nair
Priya Nair is a software developer from India. As a teenager, she created educational games to help children improve their mathematics skills. Later, she developed a website that provides free learning materials and online lessons. The platform is now used by students in several countries. Priya currently leads a team developing artificial-intelligence tools for schools. She hopes her work will reduce educational inequality worldwide.

D. Ethan Brooks
Ethan Brooks is a British musician and content creator. He started posting music tutorials online to share ideas with other musicians. His videos soon attracted millions of viewers. Ethan later created online courses for beginner musicians. He performs at festivals and charity events and is preparing for a global tour. He also plans to establish a foundation that will provide music education for disadvantaged young people.

Lexis: English Vocabulary Profile (American English)

Analysis Options

This tool assigns each word a CEFR level, according to the English Vocabulary Profile from Cambridge University Press. You can change the English type in the Analysis Options above.

If you want to correct the analysis, click on the "G" icon next to a word below, change the tag, then click Update.

Word List Word Charts

Lexis: EVP Export

See All	CEFR	Unique Words	Words	Cumul % Unique Words	Cumul % Words
	A1	70	150	37.23%	50.51%
	A2	40	49	21.28%	16.50%
	B1	33	37	17.55%	12.46%
	B2	14	16	7.45%	5.39%
	C1	6	6	3.19%	2.02%
	C2	2	2	1.06%	0.67%
	Unlisted	23	37	12.23%	12.46%

Show All

Revised Assessment Items

Questions 1–10

Read the texts about four innovators. Which person (A–D) matches each statement?

The people may be chosen more than once.

1. This person developed a project after identifying a community problem.
2. This person first became interested in their field through participation in an organised activity.
3. This person created something to improve access to education.
4. This person intends to start a larger organisation in the future.
5. This person uses their success to support younger people.
6. This person's work became widely known through online platforms.
7. This person hopes to address a global social issue.
8. This person plans to publish a major piece of work in the future.
9. This person transformed a personal interest into a successful career.
10. This person is currently collaborating with other institutions.

Answer Key

1. A
2. B
3. C
4. A
5. B
6. B
7. C
8. B
9. D
10. A

Revision Summary

Changes Made

- Reduced the length of each profile by removing supporting details and examples.
- Reduced the number of items from 12 to 10.
- Retained the four-text structure and multiple-matching task format.

Why the Changes Were Necessary

- The classroom requirement specified a shorter text.
- Shorter passages reduce reading time while maintaining sufficient information for item answering.

How CEFR Alignment Was Maintained

- The assessment remains at B2 level by requiring learners to scan multiple texts and identify information about achievements, motivations, experiences, and future plans.
- Vocabulary and grammatical complexity remain consistent with upper-intermediate reading materials.

How the Intended Cognitive Processes Were Preserved

The revised assessment continues to require learners to:

- Scan multiple texts for specific information.
- Locate relevant details rapidly.
- Identify explicit factual information.
- Recognize achievements, motivations, and future intentions.
- Compare information across texts.
- Distinguish similar but non-identical details across the four profiles.