

EXPLORING THE NEED FOR CEFR-BASED RECEPTIVE ENGLISH ASSESSMENT IN INDONESIAN JUNIOR HIGH SCHOOLS

Muhammad Rizal

rizalmuh@students.unnes.ac.id

Master's English Education Program, Faculty of Languages and Arts, Universitas
Negeri Semarang, Semarang, Indonesia

Novia Trisanti

novia.santi@mail.unnes.ac.id

Faculty of Languages and Arts, Universitas Negeri Semarang, Semarang, Indonesia

Abstract

English language teaching in Indonesian junior high schools has shifted toward a competence-based approach under Indonesia's current curriculum, whose learning outcomes are broadly comparable to CEFR A1–A2 receptive descriptors. Despite this correspondence, classroom-based instruments for assessing students' listening and reading abilities at these levels remain limited. This study explored the assessment needs of grade 9 students and English teachers prior to developing a CEFR-based receptive English assessment. Using a qualitative descriptive design, data were collected from 30 grade 9 students and two English teachers at a public junior high school in Bogor, Indonesia, through a Likert-scale questionnaire, semi-structured interviews, and analysis of nine Grade 7–9 assessment documents. Students agreed with all targeted receptive components (83.3–96.7%) and identified listening as their greatest difficulty (73.3%). The teachers were aware of the CEFR but did not apply its descriptors, largely because they lacked a practical classroom model. The assessment documents contained tasks compatible with A1–A2 descriptors but included limited listening content and no CEFR-level labels. These findings provide a blueprint for the instrument: tasks indexed to A1–A2 descriptors, a substantial listening component, items situated in familiar school and daily-life contexts, and scoring keys with level-interpretation guidance for teachers. The study contributes needs-based design evidence for CEFR assessment development at the junior secondary level, where such evidence remains limited.

Keywords: *CEFR, needs analysis, receptive English assessment, Indonesian Junior High schools*

Sari

Pengajaran bahasa Inggris di sekolah menengah pertama Indonesia telah bergeser ke pendekatan berbasis kompetensi di bawah kurikulum Indonesia saat ini, yang hasil belajarnya secara umum sebanding dengan deskriptor reseptif CEFR A1–A2. Meskipun ada korespondensi ini, instrumen berbasis kelas untuk menilai kemampuan mendengarkan dan membaca siswa pada tingkat ini masih terbatas. Studi ini mengeksplorasi kebutuhan penilaian siswa kelas 9 dan guru bahasa Inggris sebelum mengembangkan penilaian bahasa Inggris reseptif berbasis CEFR. Dengan menggunakan desain deskriptif kualitatif, data dikumpulkan dari 30 siswa kelas 9 dan dua guru bahasa Inggris di sekolah menengah pertama negeri di Bogor, Indonesia, melalui kuesioner skala Likert, wawancara semi-terstruktur, dan analisis sembilan dokumen penilaian kelas 7–9. Siswa setuju dengan semua komponen reseptif yang ditargetkan (83,3–96,7%) dan mengidentifikasi mendengarkan sebagai kesulitan terbesar mereka (73,3%). Para guru mengetahui CEFR tetapi tidak menerapkan deskriptornya, terutama karena mereka tidak memiliki model kelas yang praktis. Dokumen penilaian berisi tugas yang kompatibel dengan deskriptor A1–A2 tetapi mencakup konten mendengarkan terbatas dan tidak ada label tingkat CEFR. Temuan ini memberikan cetak biru untuk instrumen: tugas yang diindeks dengan deskriptor A1–A2, komponen mendengarkan yang substansial, item yang berada dalam konteks sekolah dan kehidupan sehari-hari yang familiar, serta kunci penilaian dengan panduan interpretasi tingkat untuk guru. Studi ini memberikan bukti desain berbasis kebutuhan untuk pengembangan penilaian CEFR di tingkat menengah pertama, di mana bukti tersebut masih terbatas.

Kata kunci: *CEFR, analisis kebutuhan, penilaian bahasa Inggris reseptif, SMP Indonesia*

Received 2026/01/12

accepted 2026/09/12

published 2026/09/30

APA Citation: Rizal, M. & Trisanti, N. (2026). Exploring the Need for CEFR-Based Receptive English Assessment in Indonesian Junior High Schools. *Research and Innovation in Language Learning*, 9(3), pp. 269-284. <http://dx.doi.org/10.33603/rill.v9i3.12681>

Introduction

Assessment is an essential part of English language teaching because it provides information that may not be fully captured through everyday classroom observation. It helps teachers understand learners' current abilities and identify areas that require further development (Bachman & Palmer, 2010). This role is particularly relevant in Indonesia, where English is widely taught as the main foreign language in formal education and continues to influence students' future academic and professional opportunities (Renandya et al., 2018; Zein et al., 2020). For this reason, understanding learners' abilities from an early stage is important. Junior high school is a particularly significant period, as students begin to develop more lasting views of their own language competence, while teachers make instructional decisions that may shape their future learning. However,

when assessment focuses primarily on previously taught materials, it may indicate how well students have mastered particular content without clearly showing their actual proficiency level or identifying what they need to develop further (Bachman & Palmer, 2010; Purpura, 2016).

The Common European Framework of Reference for Languages (CEFR) offers a different approach by describing language ability through a common scale of can-do descriptors rather than measuring achievement against specific course content (Council of Europe, 2020). For beginning learners, A1 and A2 represent the basic user levels. At these levels, learners are generally expected to understand familiar expressions, short texts, simple instructions, and routine information related to immediate personal and social situations. Receptive skills are particularly relevant at the A1–A2 levels, where learners are expected to understand clear spoken input and identify key information in simple written texts (Benigno & de Jong, 2016; Council of Europe, 2020). A level-referenced assessment of these skills can therefore offer an interpretable account of what learners are able to understand in the target language (Bachman & Palmer, 2010; Purpura, 2016).

Studies from other countries in the region indicate that CEFR adoption involves more than simply transferring or translating its descriptors into a local context. In Thailand, early implementation of CEFR-referenced assessment revealed continuing challenges in interpreting and applying the descriptors in practice (Foley, 2019). Subsequent studies have documented ongoing efforts to reconcile international proficiency descriptors with local curriculum priorities (Phoolaikao & Sukying, 2021), to develop task specifications appropriate for secondary school classrooms (Sriwongsa & Sukying, 2024), and to construct university-level tests, where the CEFR was found to be too broad to support detailed test specifications without an additional framework (Sridhanyarat et al., 2021). Comparable findings have emerged from Türkiye. Yüce and Mirici (2022) reported that CEFR-based self-assessment principles were difficult to sustain in classroom practice unless teachers were provided with clear guidance and locally applicable models. At the regional level, a systematic review of CEFR-aligned English curriculum frameworks across ASEAN countries reported a similar finding (Hamid et al., 2025). These experiences offer an important lesson for Indonesia.

Adopting CEFR labels alone is not sufficient. The descriptors must also be adapted into assessment instruments that reflect the actual conditions and needs of local classrooms.

The current policy context in Indonesia makes the adapted assessment both relevant and

necessary. Under the Merdeka Curriculum, Phase D learning outcomes expect junior high school students to understand main ideas and important details in spoken and written texts, work with familiar and unfamiliar texts that use predictable structures and familiar vocabulary, identify the purposes of texts, and begin to interpret implicit meaning (*Badan Standar, Kurikulum, dan Asesmen Pendidikan* [BSKAP], 2024). Although BSKAP does not officially assign CEFR levels to Phase D, these outcomes are broadly comparable to A1–A2 receptive descriptors in listening and reading. In this study, the relationship is therefore treated as an analytical mapping rather than an official statement of curriculum alignment. Phase D covers Grades 7 to 9 and ends in Grade 9, the final year of junior secondary education in Indonesia. Grade 9 therefore provides an appropriate point for examining students' receptive proficiency. However, curriculum expectations alone do not determine the way students are assessed. The extent to which classroom assessment reflects CEFR levels also depends on teachers, whose practices are shaped by their assessment literacy, institutional routines, and access to suitable assessment models (Aria et al., 2021; Damayanti et al., 2024; Fitriyah et al., 2022; Zulaiha et al., 2020).

Research on CEFR assessment topics in Indonesia has grown over the past decade, but most studies have focused on higher education and productive language skills. For example, Novawan et al. (2023) reported a mismatch between CEFR-based proficiency expectations for university entry and the actual proficiency levels of incoming students. This gap suggests that level-referenced assessment may need to be introduced earlier in the education system. Abi Maryo (2021) also stated that CEFR implementation should be adjusted to national curriculum priorities rather than adopted directly. Similarly, Ameliana and Zahro (2025) showed that the alignment between textbooks and CEFR-related learning outcomes cannot simply be assumed and needs to be examined systematically. A review of CEFR research trends in Indonesia further found that empirical studies remain limited, particularly those involving school age learners (Robbani et al., 2023; see also Hoinbala & Marpaung, 2025). As a result, receptive assessment at the junior high school level, particularly within the A1 to A2 proficiency range, remains largely underexplored.

This gap has clear practical implications for how the teachers deliver and evaluate the materials based on the Phase D stage. In addition, locally appropriate instruments that integrate classroom assessment with CEFR-based proficiency are still limited. This limitation can affect instructional decisions, the quality of feedback students receive about their proficiency, and the comparability of assessment results across classrooms and

schools. Before the CEFR-aligned assessment is developed, it is important to understand the needs of those who will use it. This study therefore addressed three research questions: (1) What are Grade 9 students' perceived needs regarding CEFR-based English assessment at the A1- A2 level? (2) What are English teachers' classroom assessment practices and their perceived needs regarding CEFR-based assessment? (3) To what extent do existing Grade 7- 9 English assessment documents align with CEFR A1- A2 receptive descriptors and Phase D learning outcomes? The study is intended to make a practical contribution to assessment design. Rather than simply restating the importance of the CEFR in Indonesia, it brings together evidence from students, teachers, and assessment documents to develop a set of specifications for a CEFR A1- A2 receptive assessment instrument. The development and validation of this instrument will form the next stages of the larger research project.

Method

A qualitative descriptive design was used because the study aimed to provide a clear description of what students, teachers, and existing assessment documents suggest about local assessment needs (Sandelowski, 2000). The research questions were not intended to examine cause and effect relationships or produce statistical generalisations. Instead, this study focuses on describing a particular educational setting as accurately as possible and in a way that reflects the teaching and learning experiences. Within this design, questionnaire frequencies and percentages were used only to summarise response patterns and were not treated as a basis for statistical conclusions.

The participants were 30 Grade 9 students and two English teachers from a public junior high school in Bogor, Indonesia. The decision to focus on one school was made by design rather than as a limitation that emerged after the study. This needs analysis forms the first stage of a larger development project and was intended to ground the proposed assessment instrument in the conditions of a specific and typical educational setting before broader validation is carried out. The school was selected purposively (Campbell et al., 2020) because its student population, curriculum implementation, and English learning resources broadly correspond to those of public junior high schools in the surrounding area. The site was also selected because research access had already been established for the wider project. All participants were directly involved in English teaching, learning, or assessment at the junior secondary level. The study was conducted with formal permission from the school. Written informed letter was obtained from the

participating teachers and from the students' parents or guardians, while the students also provided assent. All responses were anonymised and used only for research purposes. Data were collected from three sources to ensure that the assessment needs were not identified from a single perspective (Creswell & Creswell, 2018). The student questionnaire consisted of twenty Likert-scale items organized into five sections: previous assessment experience, perceived language difficulties, preferences for test content and format, test administration conditions, and feedback and CEFR-level reporting. Although the questionnaire consisted of 20 items across 5 sections, the findings present only the items most directly related to the research questions and the development of a CEFR A1–A2 receptive assessment instrument. A four-point response scale was used to avoid a neutral midpoint and encourage students to express a clearer position. Teacher data were gathered through semi-structured interviews conducted in Bahasa Indonesia. Each interview lasted approximately 30 to 40 minutes, was audio recorded, and transcribed verbatim. Then, relevant statements were translated into English for reporting. The document analysis examined nine English assessment documents used in Grades 7-9 during the 2024- 2025 academic year. The documents consisted of one formative assessment, one mid-semester examination, and one final semester examination from each grade level.

The questionnaire items were developed based on the CEFR A1- A2 receptive descriptors (Council of Europe, 2020) and the Phase D learning outcomes (BSKAP, 2024). Two expert validators with complementary expertise in language assessment and EFL pedagogy reviewed the items. The semi-structured interview protocol was informed by Indonesian studies on language assessment literacy and CEFR implementation (Fitriyah et al., 2022; Zulaiha et al., 2020). For the document analysis, a coding framework was developed from the same CEFR and curriculum sources. It covered five dimensions: receptive skill coverage, text types, language knowledge components, contextual familiarity, and CEFR level labelling. The first author completed the initial coding, while the second author independently reviewed the results. Differences in interpretation were discussed until agreement was reached and the final coding was established. Trustworthiness was further supported through peer debriefing with a senior researcher in language assessment and through an audit trail containing the raw data and records of coding decisions.

Questionnaire responses were calculated as percentages for each item and interpreted using a four-level descriptive classification. Percentages from 76% to 100%

were classified as strongly positive, 56% to 75% as positive, 40% to 55% as negative, and below 40% as strongly negative. Because the study was descriptive in nature, no inferential statistical procedures were applied. The interview and document data were analysed thematically through the processes of data condensation, data display, and conclusion drawing (Miles et al., 2014).

Findings and Discussion

Student questionnaire findings

The questionnaire results indicate clear student preferences for the proposed English assessment. As Table 1 shows, agreement exceeded 80% for every item related to assessment components and contextual relevance, placing all items in the strongly positive category.

Table 1. Students' perceptions of CEFR-based English assessment components

Observed items	% Agreed	Category
Importance of English assessment for measuring learning development	96.7	Strongly Positive
English assessment should help identify proficiency level	96.7	Strongly Positive
English assessment should measure listening	93.3	Strongly Positive
English assessment should measure reading comprehension	83.3	Strongly Positive
English assessment should include grammar and vocabulary	90.0	Strongly Positive
English assessment should use daily-life contexts	93.3	Strongly Positive
Multiple-choice format is easier to complete	96.7	Strongly Positive
Average across items	92.9	Strongly Positive

Two patterns in Table 1 warrant further discussion. The first is diagnostic: students want the assessment to inform them about their proficiency level, not merely provide a score, and they consider this almost as important as the assessment itself. The second concerns format and context. Agreement on daily-life contexts and the multiple-choice format was nearly as high as agreement on the skill components. This suggests that familiar content may support comprehension and that predictable formats may reduce task anxiety. Reading received the lowest level of agreement, although it remained within the strongly positive category. Figure 1 visualizes this pattern, with three items clustering at the top of the scale.

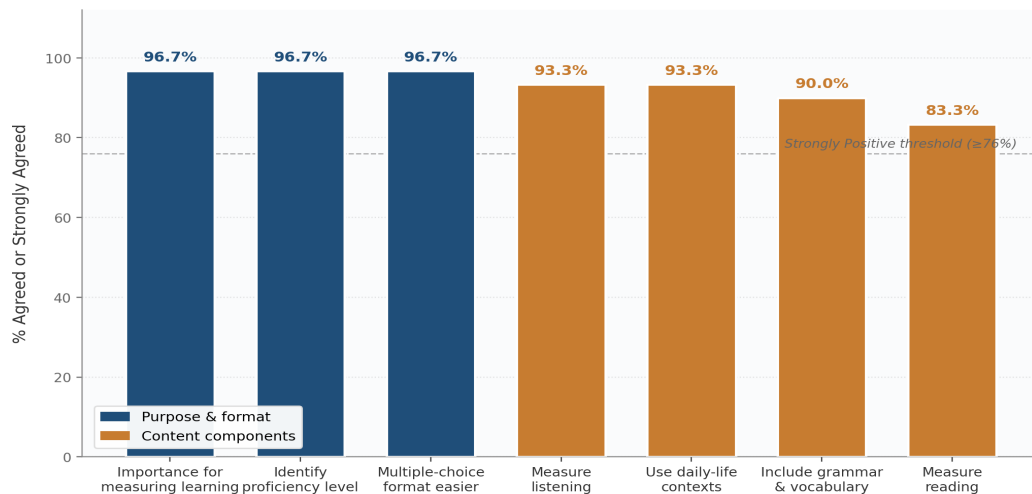


Figure 1. Students' perceptions of CEFR-based English assessment components

Table 2. Students' reported difficulties in English assessment

Observed items	% Reporting	Category
Difficulty in understanding listening questions	73.3	Positive (substantial)
Difficulty with grammar items	53.3	Negative (moderate)
Assessment anxiety	53.3	Negative (moderate)

Table 2 shifts from what students need to the difficulties they experience. Listening stands out: nearly three quarters of the students reported difficulty understanding listening items, compared with just over half who reported grammar difficulty or assessment anxiety. Read alongside Table 1, this is the questionnaire's most important finding. Students regarded listening assessment as highly important while also identifying listening as their most difficult skill.

This pattern suggests that students may value listening assessment because it addresses an area of difficulty. For instrument design, listening should therefore be treated as a central component, with careful attention to input level, audio quality, and task predictability. These findings answer the first research question by showing that Grade 9 students need a CEFR A1–A2 receptive assessment that measures listening and reading, integrates grammar and vocabulary support, uses familiar daily-life contexts, provides proficiency-level information, and gives particular attention to listening because it was identified as the most difficult skill.

Teacher interview findings

The interviews with the two teachers produced four recurring themes, summarized in

Table 3.

Table 3. Themes identified from teacher interviews

Theme	Summary of teacher responses
Reliance on formative assessment	Both teachers reported that formative assessment was the most frequently used assessment because it allowed them to monitor students' understanding after each topic and adjust subsequent instruction.
Variability of student proficiency	Teachers observed considerable variation in students' English proficiency within a single class, which required assessment tasks that could accommodate different ability levels.
Awareness but non-use of CEFR	Both teachers said having heard of CEFR through professional development activities but had not systematically applied CEFR descriptors in their classroom assessment practice.
Need for a contextualized, ready-to-use instrument	Teachers expressed openness to CEFR-based assessment provided that it was adapted to students' ability and classroom realities and accompanied by clear instructions, level indicators, and interpretation guidelines.

The four themes together provide a clear picture of classroom assessment practice. Formative assessment was used most often because it allowed the teachers to check students' understanding after each topic and make adjustments to their teaching when necessary. At the same time, the teachers reported noticeable differences in proficiency within the same class, creating a need for assessment tasks that could accommodate students with different ability levels. Although both teachers were familiar with the CEFR through professional development activities, neither had applied its descriptors into classroom assessment. As one teacher explained, "I have heard about CEFR in training, but to be honest I have not really applied it in the classroom. There is no clear guidance for how to use it with our students" (Teacher A, translated from Indonesian). Their interest in using CEFR-based assessment therefore appeared to depend on the availability of practical support. In particular, they needed an instrument that matched their students' abilities and classroom conditions, together with clear instructions, proficiency-level indicators, and guidance for interpreting the results.

This finding is consistent with previous studies in Indonesia showing that classroom assessment practices are shaped by teachers' assessment literacy, their understanding of curriculum expectations, and the conditions in which assessment is carried out (Aria et al., 2021; Damayanti et al., 2024; Fitriyah et al., 2022; Zulaiha et al., 2020). It also aligns with earlier arguments that CEFR implementation in Indonesia

requires adaptation to the local context rather than the direct application of international descriptors (Abi Maryo, 2021; Hoinbala & Marpaung, 2025; Novawan et al., 2023; Robbani et al., 2023).

In response to the second research question, the teachers' assessment practices were mainly formative and curriculum-based, while their perceived needs centred on a practical CEFR-oriented instrument with clear instructions, level indicators, and score interpretation guidance.

Document analysis findings

Table 4 summarizes the analysis of the nine Grade 7–9 assessment documents against the coding scheme derived from CEFR A1–A2 receptive descriptors and Phase D learning outcomes.

Table 4. Document analysis of Grade 7–9 English assessment materials

Dimension	Findings	CEFR/Phase D alignment
Receptive skill coverage	Reading items strongly represented; listening items limited within written assessment documents.	Partial – reading aligned; listening underrepresented.
Text types	Short functional texts, dialogues, schedules, labels, procedure texts, narratives, recounts, and reports.	Compatible with A1–A2 text-type expectations.
Language-knowledge components	Grammar and vocabulary items included across grade levels.	Compatible with A1–A2 vocabulary and structure descriptors.
Contextual familiarity	Many tasks drew on school routines and daily-life contexts familiar to learners.	Aligned with A1–A2 emphasis on familiar contexts.
CEFR-level labelling	No explicit CEFR-level labelling found in the analyzed documents.	Not aligned – implicit only.

The documents turn out to be closer to CEFR compatibility than might be expected, and further from CEFR usability. On the compatible side, the text types in use (short functional texts, dialogues, schedules, labels, procedure texts, narratives, recounts, and reports), the grammar and vocabulary coverage, and the contextual choices all sit comfortably within A1–A2 expectations. Two gaps, however, run through the whole

document set. Listening is scarcely present, so the written assessments underrepresent the very skill that students find hardest and that Phase D requires. And no document carries a CEFR-level label, so even where a task matches an A2 descriptor, neither teacher nor student can see that fact or use it when interpreting results. One example illustrates both points at once. The Grade 9 end-of-semester examination contained a recount text of roughly 120 words about a school excursion, followed by five items targeting the main idea and specific information. In response to the third research question, the existing Grade 7–9 assessment documents showed partial alignment with CEFR A1–A2 receptive descriptors and Phase D outcomes: reading-related tasks and familiar text types were broadly compatible, but listening assessment and CEFR-level labelling were still underrepresented.

Triangulation and theoretical contribution

Considered together, the three data sources point to a consistent set of assessment needs. Students support the inclusion of the targeted receptive skills and language-knowledge components, while identifying listening as the area in which they experience the greatest difficulty. Teachers recognise the value of CEFR-referenced assessment but do not yet have an instrument that can be readily integrated into their existing classroom practices. Meanwhile, the assessment documents include materials broadly compatible with A1–A2 proficiency, but they provide no listening component and do not indicate proficiency levels. This consistency across sources is methodologically significant because the questionnaire findings are supported by separate evidence from teacher interviews and authentic assessment documents, rather than being based on only one source of information.

Theoretically, the findings support a specific claim about CEFR alignment in school-level EFL contexts: descriptor translation is not enough. Test developers in the region have reached the same conclusion from the construction side, finding that CEFR descriptors must be supplemented with explicit specification frameworks before usable test items can be written (Sridhanyarat et al., 2021); the present study reaches it from the needs side. The CEFR provides a broad proficiency framework, but its descriptors are not always directly suited to the developmental and contextual characteristics of young learners and therefore require age-appropriate selection and adaptation for school-based assessment (MacGregor et al., 2025). At the A1–A2 level, this reworking involves three layers of design: connecting descriptors to national curriculum outcomes, converting

those connections into task specifications that fit classroom realities, and giving teachers interpretive guidance for using the resulting scores. The needs analysis provides evidence that all three layers are required at once. An instrument that indexed levels without classroom-fit tasks would repeat the awareness-without-use pattern found in the interviews; classroom-fit tasks without level indexing would reproduce the unlabeled compatibility found in the documents.

The findings should be interpreted within the limitations of the study. The data source is limited to one public junior high school in Bogor and involved 30 students and two teachers. For this reason, the findings are intended to inform assessment design in similar educational contexts rather than represent junior high schools across Indonesia. The questionnaire also focused on participants' perceptions of their needs and difficulties, rather than directly measuring their language performance, so the responses may be affected by some personal bias. In addition, the document analysis examined only written assessment materials and did not investigate how the assessments were carried out in actual classroom settings. These limitations will be considered in the next stages of the larger research project through instrument development, expert validation, and piloting across a broader range of school contexts.

Conclusions

This study investigated types of assessment needed for Grade 9 students and their English teachers at an Indonesian junior high school from a CEFR A1-A2 receptive assessment, along with what existing assessment documents demonstrate about current practice. All three sources of information suggest the need to use a comprehensive assessment that includes listening, reading, vocabulary, and grammar in context. Moreover, teachers need a tool that can help them distinguish among different levels and provide instruction for using it.

The existing assessment materials contain substantial reading content and many tasks that are generally compatible with A1-A2 proficiency, but there have not been listening tests or CEFR-level assessment mappings provided. In addition to these findings, the study also makes a methodological contribution to research on CEFR implementation. By examining students, teachers, and actual assessment materials together, the needs analysis provides evidence for assessment design that could not be obtained from descriptor analysis alone. Regarding the development stage, the results suggest four main dimensions to be included in the assessment framework. Firstly, the test should have

an appropriate level index. Every task in the test must be assigned to either the A1 or A2 receptive descriptor, along with the associated learning outcome at Phase D. This process must be summarized in the test blueprint. Second, the instrument should include a listening section with recorded audio that reflects A1-A2 input levels and is delivered at a pace and level of clarity appropriate for Grade 9 students. This is important because listening was the area students found most difficult and should therefore be assessed directly rather than left out. Third, the test items should use school-related and everyday situations that are familiar to Grade 9 learners, while maintaining the multiple-choice format that most participants preferred. Fourth, the assessment should be supported by a teacher package containing administration instructions, answer keys, scoring guidelines, and explanations for interpreting proficiency levels. With this assistance, the teachers will be able to apply the test and analyze its results without having further training. In the subsequent stages of this wider research project, efforts will be made to develop, validate, and pilot a testing tool that meets these requirements. The combination of student questionnaires, teacher interviews, and document analysis used in this study may also offer a useful model for conducting needs analysis in other phases of the Merdeka Curriculum.

Declaration of AI Use

During the writing of the current manuscript, the authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing, formatting, and revising the draft. During the literature search and screening, Elicit was used. The authors reviewed and verified all material that has been generated and revised from the original sources. The authors bear total responsibility for the accuracy of the content of this manuscript.

References

- Abi Maryo, F. A. (2021). The issues of the implementation of CEFR in Indonesia. *Aplinesia: Journal of Applied Linguistics Indonesia*, 5(1), 18–22. <https://doi.org/10.30595/aplinesia.v5i1.12080>
- Ameliana, A., & Zahro, S. K. (2025). Assessing the alignment of *English for Change* textbook within the Merdeka Curriculum framework. *JOLLT Journal of Languages and Language Teaching*, 13(4), 2091–2102. <https://doi.org/10.33394/jollt.v13i4.15433>
- Aria, D., Sukyadi, D., & Kurniawan, E. (2021). Teacher assessment literacy: Indonesian

- EFL secondary teachers' self-perceived on classroom-based assessment practice. *English Review: Journal of English Education*, 10(1), 15–26. <https://doi.org/10.25134/erjee.v10i1.5349>
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Benigno, V., & de Jong, J. (2016). The “Global Scale of English Learning Objectives for Young Learners”: A CEFR-based inventory of descriptors. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 43–64). Springer. https://doi.org/10.1007/978-3-319-22422-0_3
- Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2024). *Capaian pembelajaran pada pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah pada Kurikulum Merdeka*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652–661. <https://doi.org/10.1177/1744987120927206>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment—Companion volume*. Council of Europe Publishing.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Damayanti, F. Y., Setyaningsih, E., & Sumardi, S. (2024). Teachers' language assessment literacy: Indonesian secondary EFL teachers' knowledge, perspective, and classroom practice. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 13(2), 171–185. <https://doi.org/10.18592/tarbiyah.v13i2.13602>
- Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, 12(1), 124–134. <https://doi.org/10.17509/ijal.v12i1.46539>
- Foley, J. A. (2019). Issues on the initial impact of CEFR in Thailand and the region. *Indonesian Journal of Applied Linguistics*, 9(2), 359–370. <https://doi.org/10.17509/ijal.v9i2.20233>

- Hamid, H. A., Wan Fakhruddin, W. F. W., Ahmad Afip, L., & Hamid, M. O. (2025). Navigating global standards in local contexts: A systematic review of CEFR-aligned English language curriculum in the ASEAN region. *Journal of Nusantara Studies*, 10(1), 324–352. <https://doi.org/10.24200/jonus.vol10iss1pp324-352>
- Hoinbala, F. R., & Marpaung, T. (2025). Exploring the adaptability of CEFR descriptors in the construction of EFL reading test specifications in Indonesian. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 13(1), 2549–2566.
- MacGregor, D., Chapman, M., MacMillan, F., & Schultz, K. (2025). Is this even B2? Challenges of selecting CEFR descriptors for young learner assessment linking studies. *Language Teaching*, 57(4), 616–619. <https://doi.org/10.1017/S0261444824000223>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Novawan, A., Tosalem, S. M. P. A., Binarkaheni, S., & Mariana, E. R. (2023). Reflecting on the Common European Framework of Reference for Languages: Learning, teaching, and assessment (CEFR) in the Indonesian context. *Journal of English in Academic and Professional Communication*, 9(1), 1–10. <https://doi.org/10.25047/jeapco.v9i1.3754>
- 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>
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Renandya, W. A., Hamied, F. A., & Nurkamto, J. (2018). English language proficiency in Indonesia: Issues and prospects. *The Journal of Asia TEFL*, 15(3), 618–629. <https://doi.org/10.18823/asiatefl.2018.15.3.4.618>
- Robbani, A. S., Nurdianto, T., Annafiri, A. Z., Haqqy, A. M., & Irfan, I. (2023). Tren penelitian dan tantangan penerapan Common European Framework of Reference for Languages (CEFR) di Indonesia [Research trends and challenges in implementing the Common European Framework of Reference for Languages (CEFR) in Indonesia]. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 6(4), 1011–1026. <https://doi.org/10.30872/diglosia.v6i4.736>
- Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in*

- Nursing & Health*, 23(4), 334–340. [https://doi.org/10.1002/1098-240X\(200008\)23:4%3C334::AID-NUR9%3E3.0.CO;2-G](https://doi.org/10.1002/1098-240X(200008)23:4%3C334::AID-NUR9%3E3.0.CO;2-G)
- Sridhanyarat, K., Pathong, S., Suranakkharin, T., & Ammaralikit, A. (2021). The development of STEP, the CEFR-based English proficiency test. *English Language Teaching*, 14(7), 95–106. <https://doi.org/10.5539/elt.v14n7p95>
- Sriwongsa, O., & Sukying, A. (2024). Overcoming challenges: Implementing the CEFR in Thai secondary school English classrooms. *Journal of English Language Teaching and Applied Linguistics*, 6(3), 6–13. <https://doi.org/10.32996/jeltal.2024.6.3.2>
- Yüce, E., & Mirici, İ. H. (2022). Self-assessment in EFL classes of secondary education in Türkiye: The Common European Framework of Reference for Languages (CEFR)-based implementations. *Pegem Journal of Education and Instruction*, 13(1), 349–359. <https://doi.org/10.47750/pegegog.13.01.38>
- Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011–2019). *Language Teaching*, 53(4), 491–523. <https://doi.org/10.1017/S0261444820000208>
- Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An investigation into EFL teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education*, 9(1), 189–201. <https://doi.org/10.13187/ejced.2020.1.189>

Conflict of Interest

No Potential conflict of interest was reported.

About authors

Muhammad Rizal is a master's English Education Program, Faculty of Languages and Arts, Universitas Negeri Semarang, Semarang, Indonesia. He is available at rizalmuh@students.unnes.ac.id

Novia Trisanti is teaching staff at Faculty of Languages and Arts, Universitas Negeri Semarang, Semarang, Indonesia. She is available at novia.santi@mail.unnes.ac.id