p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

ENGLISH REVIEW Journal of English Education
Published by English Language Study Program, Faculty of Teachers Training & Education, Universitas Kuningan, Indonesia in collaboration with AISEE (The Association of Indonesian Scholars of English Education)
AISEE

LEARNERS' NEEDS FOR CEFR-INFORMED INTEGRATIVE ENGLISH ASSESSMENT: A MIXED-METHODS NEEDS ANALYSIS ACROSS SECONDARY AND UNIVERSITY LEVELS IN INDONESIA

Imroatul Mafiyah[1], Lubna Ali Muhammed[2], Ida Ismiyati[3], Oktiana Handini[4]
[1,3,4]*Universitas Slamet Riyadi, Indonesia*
[2]*School of Education, Lincoln University, Malaysia*

Corresponding author's email: imroatulhardiyanto@gmail.com

APA Citation: Mafiyah, I., Muhammed, L. A., Ismiyati, I., & Handini, O. (2026). Learners' needs for CEFR-informed integrative English assessment: A mixed-methods needs analysis across secondary and university levels in Indonesia. *English Review: Journal of English Education, 14*(3), 1085-1096. https://doi.org/10.25134/erjee.v14i3.356

Received: 17-06-2026 Accepted: 25-08-2026 Published: 30-10-2026

Abstract: This study aimed to identify learners' and educators' needs for CEFR-informed integrative English assessment across secondary and university levels in Indonesia as an initial stage of a larger design-based research project. Adopting a mixed-methods needs-analysis design, the study involved 105 students, comprising forty secondary and sixty-five university students. Data were collected through a questionnaire, interviews with ten students and seven English educators, supported by classroom observations and document analysis. Quantitative responses were examined using descriptive statistics, whereas qualitative evidence was thematically coded and triangulated across data sources. The results indicated that 95.2% of students regarded English proficiency as important for their academic goals, while 91.4% favoured assessment integrating reading, listening, vocabulary, and grammar. Both students and educators called for more contextualized, meaningful, and integrated assessment practices. Five principles emerged to guide future assessment design: academic relevance, functional integration, differentiated task design, contextualized linguistic resources, and transparent CEFR-informed criteria. These findings establish an empirical foundation for designing integrative assessment prototypes, which require further field testing, refinement, and validation before performance can be meaningfully interpreted in relation to CEFR levels.
Keywords: *CEFR; integrative assessment; needs analysis; EFL; higher education; secondary education.*

INTRODUCTION

English proficiency plays an important role in academic learning because students increasingly need English to access information, understand academic texts, participate in discussions, and prepare for further study. At the secondary level, English supports the development of academic literacy and prepares students for higher education.

The demands become more complex in higher education, where students encounter journal articles, textbooks, lectures, presentations, and research-related communication. Academic language use therefore rarely involves a single skill in isolation. Learners often need to move between sources and modes, for example, reading a text before discussing its ideas, listening to an explanation and taking notes, or synthesizing information from written and spoken input into a response.

These characteristics of academic communication have direct implications for language assessment. Tests that separate reading, listening, grammar, vocabulary, speaking, and writing can provide useful diagnostic information, but they do not necessarily represent the connected language activity in many academic settings (Oroh, 2025). Recent scholarship has therefore emphasized the importance of defining the target language use and clarifying what an assessment is intended to represent before task formats and score interpretations are established (Allkivi, 2026; Brunfaut, 2023; Puspitaloka et al., 2026; Tian & Crossley, 2026).

Recent research on integrated language assessment generally agrees that a test does not

become integrated simply because several language skills appear in the same examination. Integration becomes meaningful when successful completion of one part of a task depends on information processed through another language activity. Studies of reading to writing, and other source-based tasks show that learners must comprehend source material, select relevant information, reorganize it for a new purpose, and incorporate it into a new response (Harsch et al., 2024; Liu & Brantmeier, 2019; Weigle, 2004; Ye & Liu, 2023)

At the same time, recent research also shows that integrated tasks introduce more complex assessment demands. Source comprehension, source use, organization, argumentation, and linguistic control may contribute differently depending on the task and scoring criteria. Studies of rating and diagnostic frameworks therefore caution against assuming that greater integration automatically produces more valid assessment (Abbasi Sardari et al., 2025; Phakiti & Leung, 2024; Plakans et al., 2019). These studies shift the discussion from whether skills should be combined to how task relationships and scoring can represent academic language use without introducing unintended difficulty.

In Asian EFL contexts, Maryana et al. (2026) found meaningful relationships between receptive and productive performance in integrated tasks among Indonesian secondary-school learners and showed that students reused information from input when constructing later responses. Raymundo (2023) similarly reported the value of linked macro-skill assessment in Philippine higher education, while Sadeghi and Bakhshi (2025) showed that performance on integrated listening-speaking tasks may also be influenced by cognitive, metacognitive, and affective factors.

A second research strand concerns CEFR-informed assessment. The CEFR Companion Volume presents language ability in terms of purposeful reception, production, interaction, and mediation rather than isolated knowledge of forms (Council of Europe, 2020). North et al. (2022) further argue that the framework should not be reduced to level labels, because its action-oriented perspective emphasizes what learners can accomplish with language.

Recent studies highlight the need for careful contextual interpretation when applying CEFR to assessment. Piamsai (2023) reported only a moderate relationship between CEFR-based self-assessment and measured proficiency, while Balleras et al. (2025) demonstrated that test alignment requires explicit familiarization and specification procedures. Waluyo et al. (2024) likewise showed that CEFR-based university test development becomes defensible only when descriptors are accompanied by systematic analysis of item functioning, reliability, and validity.

These findings suggest that assessment innovation cannot be treated solely as a technical matter of task construction. Its feasibility and relevance are also influenced by the educational context in which assessment is implemented. This concern is consistent with Hoinbala & Marpaung (2025) who emphasize the importance of adapting assessment specifications to local curricular and instructional contexts. Accordingly, the development of more integrated forms of English assessment in Indonesia needs to consider both the target language abilities and the realities of classroom and institutional practice (Sari, 2024; Maryana et al., 2026).

Local assessment conditions also shape what can be implemented. In Indonesia, teachers have reported practical challenges involving resources, learner diversity, professional preparation, and institutional expectations (Sari, 2024). More broadly, recent work on language assessment literacy shows that teachers' assessment decisions are influenced by training, professional context, digital demands, and institutional constraints (Allkivi, 2026; Busby & Skjelde, 2026; Puspitaloka et al., 2026) Indonesian studies similarly report gaps between assessment knowledge and enacted practice (Puspawati & Murtiningsih, 2025; Sukarno et al., 2024; Umam et al., 2023). Assessment development must therefore respond not only to target language abilities but also to the educational conditions in which assessment is implemented.

A third strand of research highlights the role of needs analysis in defining that environment. Recent EAP and ESP studies show that academic language needs vary across educational stages, disciplines, and institutional purposes (Cheng, 2021; Gaffas, 2019; Tzoannopoulou, 2015) .

Large-scale and multi-source needs analyses have been used to inform EAP curriculum development (Hyland & Jiang, 2021), synthesis-writing instruction (Plakans et al., 2019), and discipline-specific materials Indonesian studies further demonstrate the value of needs analysis for identifying context-specific academic and professional language demand (Amirbayeva in English, 2024; Mella Narolita & Aulia Restu Ariyanto Putri, 2025; Sankar & Aravind, 2026) .

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

These findings support the use of multiple perspectives when defining what learners are expected to do with English.

For assessment development, however, needs analysis has a more specific function than documenting learner preferences. An argument-based view of validity treats validation as the evaluation of evidence supporting intended interpretations and uses of assessment performance (Chapelle & Voss, 2021; Kane, 2013; Lauwaert, 2023). From this perspective, assessment development begins with a defensible definition of the target domain and a clear account of the performances that should be represented. Needs analysis can contribute early evidence for domain definition and task specification, but it cannot establish score validity or CEFR alignment by itself. Such claims require additional evidence from task performance, scoring, reliability, standard setting, and validation (Chapelle & Voss, 2021; Kane, 2013)

This framing makes it possible to identify the specific gap addressed by the present study. Existing work has established that integrated tasks can reveal relationships between receptive and productive performance, and CEFR descriptors can support test specification when they are contextually interpreted. Yet these strands have largely addressed different stages of assessment development. Maryana et al. (2026) examined integrated performance at one educational level rather than the cross-level needs-analysis stage. Waluyo et al. (2024) demonstrated CEFR-based test development and validation but did not begin from a multi-source analysis of learner and educator needs across educational levels. Raymundo (2023) examined integrated-assessment practices but did not translate learner, educator, classroom, and documentary evidence into CEFR-informed design requirements. What therefore remains unresolved is how these forms of evidence can be connected at the initial assessment-development stage across secondary and university contexts.

The contribution of the present study lies in the combination of three elements. First, it adopts a cross-level perspective to identify both shared principles and differences in academic language demands between secondary and university learners. Second, it treats CEFR descriptors as design references rather than assumed proficiency classifications, separating early task specification from later claims of alignment. Third, it translates triangulated needs-analysis evidence into explicit design principles during the first phase of a Design-Based Research project. The contribution is therefore not the use of multiple data sources in itself, but the way those sources are connected to theoretically informed assessment-design decisions.

The study consequently contributes at three levels. Methodologically, it illustrates how stakeholder and contextual evidence can be used to define an assessment domain before prototype construction. Theoretically, it connects integrated-assessment research, the CEFR action-oriented view of language use, and argument-based validity by positioning needs analysis as early evidence for domain definition and task specification. Practically, it produces level-sensitive design requirements for Indonesian secondary and university settings and highlights the assessment-literacy support required for implementation (Hao & Johnson, 2013; Mohamad Marzaini et al., 2023a; Sumardi & Guci, 2023).

The study represents the needs-analysis phase of a broader Design-Based Research (DBR) project. DBR is suitable for this purpose because it begins with an educational problem situated in practice and develops solutions through iterative stages of analysis, design, implementation, evaluation, and refinement (Akker et al., 2020; Valentine & Kopcha, 2013) . At this stage, the study does not claim that a final CEFR-informed integrative assessment has already been developed or validated. Instead, the findings are intended to provide an empirical foundation for subsequent prototype development, field testing, refinement, and validation. Accordingly, the study addresses the following research questions: (1) What English skills and linguistic resources do secondary and university learners perceive as important for academic learning, and how do these perceptions differ across educational levels? (2) How do students and educators perceive current English assessment practices and the use of more integrative forms of assessment? (3) What CEFR-informed design principles can be derived from the triangulated needs-analysis evidence to guide subsequent assessment prototype development?

METHOD

This study employed an explanatory sequential mixed-methods needs analysis design within the analysis and exploration phase of a broader Design-based Research (DBR) project. In an explanatory sequential design, quantitative data are collected and analysed first, followed by qualitative inquiry that helps explain and contextualize the initial quantitative findings, with

both forms of evidence integrated during interpretation. The quantitative process was conducted first to identify general patterns in learners' perceived academic English and assessment expectations (Creswell & Creswell, 2018). The qualitative phase subsequently explored and explained these patterns through student and educator interviews, classroom observations, and document analysis.

A mixed-methods approach was selected because questionnaire data could identify general patterns in learners' perceived needs but could not fully explain why these patterns emerged or how they related to existing teaching and assessment practices. Interviews provided further explanation, while classroom observations and document analysis offered contextual evidence. The quantitative and qualitative phases were connected by using questionnaire findings to guide the interviews and by comparing evidence from questionnaires, interviews, observations, and documents during interpretation. This process helped identify shared patterns, additional explanations, and differences across participants and contexts, which were then translated into initial requirements for prototype development (Plomp & Nieveen, 2013). The study was conducted in Surakarta, Indonesia, and involved participants from secondary and university educational contexts. The quantitative sample consisted of 105 students, comprising 40 secondary school students and 65 undergraduate students. Secondary participants were approximately 15-18 years old, while university participants were 18-23 years old. The two educational groups were included to identify both shared and level-specific academic English assessment needs. The qualitative phase involved ten student interviewees representing the two educational contexts and seven English educators, consisting of four secondary-school English teachers and three university lecturers. The educators contributed professional perspectives on current assessment practices, practical classroom challenges, and the academic language demands experienced at their respective educational levels.

A purposive sampling strategy was used because the research required participants with direct experience of English learning, teaching, and assessment in the educational context involved. Student interviewees were selected to represent both educational levels and to provide more detailed accounts of issues emerging from the questionnaire. The qualitative sample was intended to provide explanatory depth rather than statistical generalization.

Table 1. *Participant profile and role in the needs analysis*

Participant Group	Educational Context	Approximate Age	Total (n)	le in the Study
Secondary students	SMA/SMK	15–18	40	Academic English needs and assessment experiences
University students	Undergraduate	18–23	65	Academic language demands and assessment experiences
Student interviewees	Secondary and university	Corresponding age groups	10	Explanation of questionnaire results
English teachers	Secondary education	30–50	4	Pedagogical and assessment perspectives
University lecturers	Higher education	35–55	3	Academic and assessment perspectives

The study used four data sources: questionnaire, semi-structured interviews, classroom observations, and document analysis. The questionnaire examined the importance of English for academic purposes, reading and listening, grammar and vocabulary, and preferences for more integrative assessment. Sample items included,

"English proficiency is important for achieving academic goals,"

"Grammar mastery is important for academic writing," and

"Tests combining reading, listening, vocabulary, and grammar provide a better evaluation of English ability."

The questionnaire was used as a needs-analysis instrument rather than a proficiency test. The questionnaire items were reviewed to ensure alignment with the research questions and study domains, as well as clarity and relevance for the participants. Internal consistency was examined using Cronbach's alpha based on the retained respondent-level responses to eight common

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

questionnaire items. The analysis yielded Cronbach's α = .937, indicating high internal consistency among the items.

Semi-structured student interviews explored experiences with existing tests, academic language activities, learning difficulties, and views of more integrated assessment. Educator interviews focused on current assessment formats, language-skill integration, scoring practices, and practical assessment needs. Example questions included "How well do your current English tests reflect the way you use English in your studies?" and "What factors should be considered when developing more integrated assessment tasks?"

Classroom observations examined the language skills used in learning activities, the treatment of grammar and vocabulary, and whether language components were used separately or together. Document analysis included existing assessment materials, curriculum or syllabus documents, and relevant CEFR descriptors. Documents were compared in terms of task format, language components assessed, degree of integration, and expected learner responses.

Data collection was conducted during the needs-analysis stage of a broader Design-Based Research project. The broader project followed an iterative sequence of practical problem analysis, prototype development, testing and revision, and reflection and refinement (Akker et al., 2020; Anderson & Shattuck, 2012; Barab & Squire, 2004). In the present study, the questionnaire was administered first to identify general patterns in learners' academic English and assessment needs. Semi-structured interviews were then conducted with selected students and English educators to explore these patterns in greater depth. Classroom observations and document analysis provided additional evidence about current teaching and assessment practices. The findings from all sources were subsequently compared and integrated to identify common patterns and formulate design requirements for subsequent prototype development. The procedures, data collection techniques, analytical methods, and contribution of each data source are summarized in Table 2.

Table 2. *Data collection and analysis*

Source of Data	Participants	Analysis	Contribution
Questionnaire	40 secondary and 65 university students	Frequencies and percentages	Identified general patterns of academic English needs and assessment preferences
Student interviews	10 students	Thematic coding and constant comparison	Explained and contextualized questionnaire patterns
Educator interviews	4 teachers and 3 lecturers	Thematic coding and constant comparison	Identified current practices, constraints, and future assessment needs
Classroom observations	Secondary and university contexts	Descriptive coding and comparison	Provided contextual evidence of actual language use and assessment practices
Document analysis	Tests, curriculum/syllabus documents, CEFR descriptors	Content comparison	Identified gaps among assessment practice, curricular expectations, and CEFR-informed design
Integration	All evidence sources	Triangulation and cross-source comparison	Generated provisional design principles for prototype development

Questionnaire responses from the 105 participants were analysed using frequencies and percentages to identify the relative salience of students' academic English and assessment needs. The archived results showed consistently high levels of agreement across key domains, including English proficiency for academic goals, academic reading and listening, grammar for academic writing, vocabulary for understanding texts, and more integrative assessment.

Interview, classroom-observation, and documentary evidence was analysed using an iterative coding and constant-comparison procedure. The first stage involved identifying recurrent ideas concerning academic language needs, learner difficulties, linguistic resources, current assessment practices, and integrated language use. These initial codes were subsequently grouped into broader analytical categories. The categories were then compared

across student and educator groups and across secondary and university contexts. Observation and documentary evidence were used to determine whether patterns reported in interviews were supported, qualified, or contradicted by contextual evidence. The final analytical stage translated patterns supported by more than one evidence source into provisional assessment-design principles. Because the qualitative evidence was intended to explain and contextualize the quantitative patterns rather than estimate prevalence, themes were interpreted according to their explanatory relevance and cross-source support rather than converted into numerical frequencies. Trustworthiness was strengthened through comparison across student interviews, educator accounts, classroom observations, and documentary evidence.

Participation was voluntary, and informed consent and confidentiality were maintained throughout the study. Participants were informed about the purpose of the research and the academic use of the data, and identifying information was excluded from reported quotations. For secondary-school participants, participation followed the permission and consent procedures applicable to the participating institutions.

RESULTS AND DISCUSSION

Academic English needs and linguistic Resources

The questionnaire findings showed consistently strong recognition of English as an important academic resource. Of the 105 participants, 95.2% agreed that English proficiency was important for achieving their academic goals. This qualitative evidence helped explain this high level of agreement. Secondary learners generally associated English with curriculum-based learning, access to information, and preparation for higher education, whereas university students described more specialized demands, including reading journal articles, following academic lectures, understanding disciplinary information, and participating in academic settings. The main difference between the two groups, therefore, concerned how and for what purposes English was used, rather than whether it was considered important. Secondary learners were still developing broader academic literacy, while university students increasingly encountered specialized texts and more demanding academic communication.

This pattern is consistent with recent needs-analysis research showing that academic English requirements vary according to educational level, disciplinary context, and target academic activities. Çiftçi B and Yavuz (2025), for example, identified differences in EAP needs across educational levels and fields, while Nausa et al. (2024) demonstrated that academic reading needs become more specific when students engage with disciplinary texts at postgraduate level. Damayanti et al. (2024) similarly showed in the Indonesian context that needs analysis can identify differences in academic language requirements and provide a basis for context-sensitive educational design.

The present findings extend this evidence by examining secondary and university learners within the same assessment-development framework. Although the two groups shared a broad need for academic English, they differed in the complexity of texts, language resources, and responses expected of them. This difference suggests that assessment should maintain common principles across educational levels while adjusting input complexity, task support, and response requirements to reflect learners' academic contexts.

Grammar and vocabulary also remained important across both groups. 93.3% of participants considered grammar important for accurate academic writing, while 91.4% regarded vocabulary as important for understanding academic texts. The qualitative findings added a cross-level distinction: university learners referred more frequently to academic vocabulary and complex grammatical structures, whereas secondary learners experienced greater difficulty with general vocabulary and basic-to-intermediate grammatical patterns.

These results suggest that linguistic resources should not be treated as irrelevant simply because the assessment aims to become more communicative or integrated. Instead, their function changes as academic demands increase. Vocabulary supports comprehension and source interpretation, while grammatical resources contribute to learners' ability to explain, summarize, and communicate information accurately. The findings therefore qualify a simple contrast between discrete linguistic assessment and integrated performance: learners may value grammar and vocabulary while also expecting them to be assessed within more meaningful contexts.

This interpretation becomes particularly important when the findings are considered alongside current assessment practices. University participants reported that grammar and vocabulary

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

were often assessed separately even though their coursework required them to read, interpret, summarize, and communicate information. In the university group, 76% indicated that examinations still concentrated mainly on grammar and vocabulary, with relatively limited opportunity to combine information or demonstrate broader academic reasoning.

Recent research on classroom assessment provides a useful explanation for this mismatch. Latif and Wasim (2024) found that traditional assessment practices may remain dominant even when educators recognize the value of broader constructivist approaches, reflecting tensions between assessment beliefs and enacted classroom practice. Similarly, contemporary language-assessment-literacy research emphasizes that assessment decisions depend not only on teachers' knowledge but also on training, institutional conditions, and the ability to translate assessment principles into classroom practice (Al-Akbari et al., 2025; Estaji, 2024; Mizoguchi, 2024).

Thus, students' strong support for grammar and vocabulary should not be interpreted as a preference for maintaining predominantly discrete testing. Rather, the quantitative and qualitative evidence together indicate that learners distinguish between the importance of linguistic knowledge and the way that knowledge is assessed. Grammar and vocabulary remain valuable, but learners expect these resources to contribute to meaningful academic language use.

Current assessment practices and perceptions of integrative assessment

The second research question focused on participants' views of current assessment practices and more integrative alternatives. Across the qualitative evidence, assessment was generally described as organized around separate skills or language components.

University educators reported that reading and listening were commonly assessed through multiple-choice or short-answer formats, while speaking and writing were assessed using different criteria. Although such arrangements were considered practical and familiar, educators questioned whether they adequately represented students' ability to use English across connected academic activities.

Student accounts strengthened this interpretation. The finding that 76% of university respondents perceived examinations as largely focused on grammar and vocabulary became more meaningful when considered alongside their descriptions of coursework. Students were expected to read academic texts, summarize information, respond to lectures, and present ideas during learning activities, but examinations did not always require the same forms of language use. The quantitative evidence therefore identifies the extent of the perceived mismatch, while the interviews explain why students felt that the existing assessment did not fully reflect their academic language needs.

Support for more integrated assessment was correspondingly high. 91.4% of participants supported assessment combining reading, listening, vocabulary, and grammar, while 89.5% believed that a more integrative approach would better represent actual language use. Interviews clarified what participants meant by such assessment.

University students referred to activities such as summarizing academic texts, responding to information from short lectures, interpreting information presented in graphs or tables, while secondary learners generally preferred shorter and more guided contextual tasks.

Previous research on integrated assessment helps interpret these preferences. Ye and Liu (2023) show that integrated writing requires learners to draw on reading comprehension while planning and producing a written response. Harsch et al. (2024) further demonstrate that integrated reading-to-writing assessment requires careful validation of the scoring scale because raters need to evaluate performance across several connected dimensions.

These studies support the present finding that integration should be functional rather than merely structural. A test containing separate reading, listening, vocabulary, and grammar sections may still operate as a collection of discrete measures. Integration becomes more meaningful when information obtained through one language activity contributes directly to another response, for example when learners read or listen to information before summarizing, comparing, explaining, or responding to it.

However, the present findings also suggest that greater integration should not automatically be regarded as preferable. Integrated tasks may place additional demands on learners because they require them to process source information, organize ideas, and produce a response while maintaining appropriate linguistic control. Abbasi Sardari et al. (2025) similarly show that integrated performance involves several interrelated dimensions that need to be clearly represented in

diagnostic and scoring criteria. Integrated assessment therefore needs to designed carefully so that task demands remain consistent with the abilities the assessment is intended to measure.

For this study, this integration should be adapted to learners' levels. Secondary learners may require shorter texts, familiar topics, clearer guidance, and more controlled responses, whereas university learners can be expected to process denser academic information and produce more independent summaries, comparisons, or explanations.

Document analysis reinforced this point. Relevant CEFR descriptors provide examples of purposeful activities such as summarizing, explaining, processing information, and delivering structured responses, yet such descriptors had not been systematically incorporated into assessment practices in the contexts examined. The CEFR may therefore provide useful reference points for task and scoring development, provided that descriptors are adapted to the local context rather than applied mechanically. This provides a basis for translating the needs-analysis into CEFR-informed design principles for further prototype development.

CEFR-informed design principles for prototype development

Integrated evidence from the needs analysis provided a basis for defining how the assessment prototype should be developed. At this stage, the purpose was not to produce a finished CEFR-level test, but to translate the needs identified across questionnaires, interviews, classroom observations, and documents into a set of evidence-based design requirements. This reflects the role of needs analysis in identifying relevant target language use and guiding the next stage of assessment development (Plomp & Nieveen, 2013; Valentine & Kopcha, 2013). Five related principles emerged from the findings: academic relevance, functional integration, differentiated task design, contextualized linguistic resources, and clear CEFR-informed criteria.

One key principle is academic relevance. The prototype should reflect language activities that are meaningful in learners' educational contexts. University learners were expected to work with academic texts, lectures, summaries, presentations, and research-related information, while secondary learners mainly needed English for curriculum-based learning and preparation for further study. Assessment tasks should therefore reflect these different purposes rather than focus only on isolated language forms.

The second principle is functional integration. The findings showed that reading, listening, grammar, and vocabulary were often assessed separately, even though learners needed to use them together in academic activities. The prototype should therefore connect input and response within the same task. Relevant tasks may include activities such as reading or listening to information and then using it to identify, compare, explain, or summarize key ideas. The literature on integrated-assessment emphasizes that integration depends on how language activities are connected within a task, rather than the number of skills included (Phakiti & Leung, 2024; Rao, 2018; Raymundo, 2023; Sadeghi & Bakhshi, 2025).

A third principle is differentiated-task design across educational levels. The same assessment should not be used in exactly the same way for secondary and university learners. In the broader research blueprint, A2–B1 and B2–C1 were used as intended reference ranges for secondary and university prototype development. Secondary-level tasks were designed around more accessible texts, familiar listening input, main ideas, and basic grammatical control, whereas university-level tasks involved academic texts, lectures, argument, inference, and more complex grammatical structures. These ranges are design references rather than confirmed CEFR levels of the participants. In practice, this means that task difficulty should be adjusted through input length, vocabulary level, information density, response complexity, and the amount of support provided. A secondary task may use a short text or listening passage followed by a guided response, while a university task may require learners to process more complex information before summarizing or comparing ideas. The CEFR can support this process by providing descriptors that help define expected performance at different levels.

The fourth principle concerns contextualized linguistic resources. Grammar and vocabulary should remain visible in the prototype because students considered both important for academic learning. However, the findings indicate that these resources should not appear only as isolated test items. At secondary level, relevant areas include basic grammar and general vocabulary, whereas university learners require more academic vocabulary and more complex grammatical resources. The prototype may still use direct grammar or vocabulary items when diagnostic information is needed, but these resources should also be embedded in reading, listening, and

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

response tasks so that learners demonstrate how they use linguistic knowledge in context.

A final design consideration involves performance expectations and scoring criteria. Students and educators indicated a need for clearer and more transparent assessment expectations, particularly in productive tasks. The report shows that existing rubrics varied and were largely developed for specific institutional contexts, while students wanted clearer criteria for understanding the quality of their performance. Document analysis also showed that relevant CEFR descriptors were available for academic activities, including summarizing information from multiple sources and producing structured presentations, but had not yet been systematically incorporated into assessment in the research context. Relevant can-do descriptors can therefore inform task expectations and scoring criteria, while later piloting and validation remain necessary before results are linked to specific CEFR levels. The relationship between the findings and these five design principles is summarized in Table 3.

Table 3. *Evidence-based design principles for prototype development*

Design Principle	Main Evidence	Implication for Prototype Development
Academic relevance	Learners needed English for educational activities that differed across secondary and university contexts	Use meaningful, level-appropriate academic or curriculum-related tasks
Functional integration	Students and educators identified a need to connect language components rather than assess them only separately	Link reading/listening input with a meaningful subsequent response
Differentiated task design	Secondary learners showed greater difficulty with listening, while university learners faced more complex academic vocabulary and grammar demands	Adjust input, vocabulary, task complexity, response demands, and support
Contextualized linguistic resources	Grammar and vocabulary remained important but were often tested as isolated components	Assess linguistic resources both directly when needed and within meaningful contexts
Clear and CEFR informed criteria	Students wanted clearer expectations; existing assessment had not systematically used relevant CEFR descriptors	Use selected can-do descriptors to inform task specifications and scoring criteria

CONCLUSION

The study addressed three questions concerning academic English needs, current assessment practices, and CEFR-informed design requirements. Regarding the first question, learners at both educational levels regarded English, reading, listening, grammar, and vocabulary as important for academic learning, although university learners faced more specialized and complex demands. Regarding the second, students and educators identified a mismatch between largely component-based assessment practices and the connected language use expected in academic learning, which helps explain the strong support for more contextualized and integrative assessment. Regarding the third, the triangulated evidence produced five design principles: academic relevance, functional integration, differentiated task design, contextualized linguistic resources, and clear CEFR-informed criteria.

The central contribution of the study is to demonstrate how cross-level, multi-source needs evidence can be translated into theoretically grounded assessment-design requirements before claims of CEFR alignment or score validity are made. Methodologically, the study shows how learner, educator, classroom, and documentary evidence can contribute to domain definition at the start of test development. Theoretically, it connects the CEFR action-oriented view of language use, integrated-assessment research, and an argument-based perspective on validity. Practically, it provides a transparent set of level-sensitive requirements for the next stage of prototype development.

Several limitations remain important. The study was conducted in specific secondary and university contexts, the CEFR ranges were developmental references rather than measured proficiency levels, and no prototype was implemented in the present

phase. The findings therefore cannot establish reliability, validity, fairness, practicality, or empirical CEFR alignment of a future assessment. They also do not demonstrate that more integrative assessment will automatically improve learning outcomes.

Future research should develop and refine prototype tasks, conduct cognitive and expert review, investigate task difficulty and scoring reliability, and examine how selected CEFR descriptors function in actual performance. Standard-setting and validation studies will be necessary before scores are linked to specific CEFR levels. Later DBR cycles should also investigate practicality, teacher assessment literacy, learner engagement, and the impact of the assessment on teaching and learning. In this way, the present needs analysis provides a defensible starting point for assessment development rather than a premature claim of a finished CEFR-aligned test.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Slamet Riyadi University for supporting this research through its internal research funding scheme. The authors also thank the participating students, teachers, and lecturers for their contributions to the needs-analysis process.

REFERENCES

Abbasi Sardari, S., Ahmadi, A., & Ravand, H. (2025). Creating an empirically derived checklist to diagnose second language learners' integrated writing ability. *Language Testing in Asia*, *15*(1). https://doi.org/10.1186/s40468-025-00396-3

Akker, J. Van Den, Gravemeijer, K., McKenney, S., & Nieveen, N. (2020). Educational Design Research. In *Educational Design Research*. Routledge. https://doi.org/10.4324/9780203088364-9

Al-Akbari, S., Nikolov, M., & Hódi, Á. (2025). EFL teachers' language assessment literacy training needs. *Social Sciences and Humanities Open*, *11*. https://doi.org/10.1016/j.ssaho.2024.101254

Allkivi, K. (2026). Towards interpretable models for language proficiency assessment: Predicting the CEFR level of Estonian learner texts. *Journal of Responsible Technology*, *25*. https://doi.org/10.1016/j.jrt.2026.100162

Amirbayeva in English, D. (2024). Needs Analysis of English for Specific Purposes (ESP) in the Tourism Industry of Indonesia and Kazakhstan. In *LEARN Journal: Language Education and Acquisition Research Network.*

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Balleras, J., Yippikun, C., Castro, R., & Prasertsak, A. (2025). Developing a Standardized English Proficiency Test Based on the CEFR and Language Test Development Manual. *Journal of English Language and Linguistics*, *6*(2), 169–189. https://doi.org/10.62819/jel.2025.1007

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, *40*(1), 15–23. https://doi.org/10.1177/02655322221127896

Busby, N. L., & Skjelde, K. M. (2026). L2 English teachers' conceptualization of academic language – where does it come from and how does it relate to teaching practices? *Journal of English for Academic Purposes*, 83. https://doi.org/10.1016/j.jeap.2026.101756

Chapelle, C. A., & Voss, E. (2021). Introduction to validity argument in language testing and assessment. In *Validity Argument in Language Testing: Case Studies of Validation Research* (pp. 1–16). Cambridge University Press. https://doi.org/10.1017/9781108669849.002

Cheng, A. (2021). The place of language in the theoretical tenets, textbooks, and classroom practices in the ESP genre-based approach to teaching writing. *English for Specific Purposes*, *64*, 26–36. https://doi.org/10.1016/j.esp.2021.07.001

Çiftçi B, E., & Yavuz, A. (2025). Investigating the Needs of English for Academic Purposes (EAP) Learners. *Innovational Research in ELT*, *6*(1), 2025. http://orcid.org/0000-0001-6838-8695

Creswell & Creswell. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publication.

Damayanti, I. L., Derinalp, P., Asyifa, F., & Suryatama, K. (2024). Exploring English for academic purposes program: needs analysis and impact evaluation. *Studies in English Language and Education*, *11*(3), 1616–1635. https://doi.org/10.24815/siele.v11i3.38329

Estaji, M. (2024). Unveiling the impact of language assessment teacher education on EFL teachers' assessment agency and professional development trajectory: perceptions, opportunities, and challenges. *Language Testing in Asia*, *14*(1). https://doi.org/10.1186/s40468-024-00292-2

Sari, R.F. (2024). English language assessment in Indonesia secondary high school : implementation and challenges. *Aplinesia (Journal of Applied Linguistics Indonesia)*, *8*(1),

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 3, October 2026

69–75. https://doi.org/10.30595/aplinesia.v8i1.29825

Gaffas, Z. M. (2019). Students' perceptions of the impact of EGP and ESP courses on their English language development: Voices from Saudi Arabia. *Journal of English for Academic Purposes, 42*, 100797. https://doi.org/10.1016/j.jeap.2019.100797

Hao, S., & Johnson, R. L. (2013). Teachers' classroom assessment practices and fourth-graders' reading literacy achievements: An international study. *Teaching and Teacher Education, 29*(1), 53–63. https://doi.org/10.1016/j.tate.2012.08.010

Harsch, C., Koval, V., Kanistra, P. (Voula), & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing, 62*. https://doi.org/10.1016/j.asw.2024.100894

Hyland, K., & Jiang, F. (Kevin). (2021). Delivering relevance: The emergence of ESP as a discipline. *English for Specific Purposes, 64*, 13–25. https://doi.org/10.1016/j.esp.2021.06.002

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement, 50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Latif, M. W., & Wasim, A. (2024). Investigating EFL instructors' approaches to classroom-based assessment culture: an explanatory sequential mixed-method approach. *Language Testing in Asia, 14*(1). https://doi.org/10.1186/s40468-024-00332-x

Lauwaert, P. (2023). On validity. *Studies in Applied Linguistics & TESOL at Teachers Colleg, 23*(1), 18–36.

Liu, H., & Brantmeier, C. (2019). "I know English": Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. *System, 80*, 60–72. https://doi.org/10.1016/j.system.2018.10.013

Maryana, H., Yusuf, F. N., & Amalia, L. L. (2026). Tracing input use across integrated EFL assessment: Relationships between receptive and productive performance. *Voices of English Language Education Society, 10*(2). https://doi.org/10.29408/veles.v10i2.34796

Mella Narolita, & Aulia Restu Ariyanto Putri. (2025). Needs analysis on english language use in tour guiding. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan, 3*(4), 5558–5565. https://doi.org/10.31004/jerkin.v3i4.1238

Mizoguchi, R. (2024). A language assessment literacy survey of pre-service and in-service teachers: A mixed-methods replication study of vogt and tsagari (2014). *JLTA Journal, 27*, 61–80.

Mohamad Marzaini, A. F., Haslee Sharil, W. N. E., Supramaniam, K., & Mat Yusoff, S. (2023). Evaluating teachers' assessment literacy in enacting Cefr-aligned classroom-based assessment in Malaysian secondary schools ESL classroom. *International Journal of Academic Research in Progressive Education and Development, 12*(1). https://doi.org/10.6007/ijarped/v12-i1/15691

Nausa, R., Živković, J., & Sichko, L. (2024). Needs analysis and design of a master's level academic reading course in English. *Profile: Issues in Teachers' Professional Development, 26*(2), 89–106. https://doi.org/10.15446/profile.v26n2.110905

North, Brian., Piccardo, Enrica., Goodier, Tim., Fasoglio, Daniela., Margonis-Pasinetti, Rosanna., & Rüschoff, Bernd. (2022). *Enriching 21st-century language education: the CEFR companion volume in practice.* Council of Europe Publishing.

Oroh Zuska Elizabeth. (2025). *Rethinking Assessment in ELT: Traditional Assessment vs Authentic Assessment*. Tahta Media Group.

Phakiti, A., & Leung, C. (2024). Assessment for language teaching. In *Assessment for Language Teaching*. Cambridge University Press. https://doi.org/10.1017/9781108934091

Piamsai, C. (2023). Development and use of CEFR based self-assessment in a Thai tertiary context (Vol. 66). https://digital.car.chula.ac.th/pasaa

Plakans, L., Liao, J. T., & Wang, F. (2019). "I should summarize this whole paragraph": Shared processes of reading and writing in iterative integrated assessment tasks. *Assessing Writing, 40*(February), 14–26. https://doi.org/10.1016/j.asw.2019.03.003

Plomp, Tj., & Nieveen, Nienke. (2013). *Educational design research*. Part A : an introduction. SLO.

Puspitaloka, N., Ermanto, & Zainil, Y. (2026). Claimed vs. measured CEFR levels: Profiling grade 8 EFL reading texts in Indonesia. social sciences & humanities open, 14, 103687. https://doi.org/10.1016/j.ssaho.2026.103687

Rao, P. S. (2018). Evaluation and assessment in ELT: A global perspective. *Research Journal of English Language and Literature (RJELAL), 6*(2). https://doi.org/10.33329/rjelal.62.237

Raymundo, J. L. (2023). Exploring the practices on macro skills integrated assessment in Philippine higher education context: Basis in designing a language training program. *International Journal of Language Education, 7*(3), 429–446. https://doi.org/10.26858/ijole.v7i3.24252

Hoinbala, F.R, Marpaung, T. (2025). The adaptability of CEFR descriptors in the construction of EFL reading test specifications in Indonesian. *IDEAS Journal of Language Teaching and Learning, Linguistics and Literature Exploring*. *13*(1), 2549–2566. https://doi.org/10.24256/ideas.v13i1.6451

Sadeghi, K., & Bakhshi, N. (2025). Integrated listening/speaking skill assessment: The role of ambiguity tolerance, cognitive/metacognitive strategy use, and foreign language anxiety.

International Journal of Language Testing, *15*(1), 77–90. https://doi.org/10.22034/ijlt.2024.452581.1338

Sankar, S. S., & Aravind, B. R. (2026). English for specific purposes in hotel and tourism management: A need analysis. *International Journal of Language and Literary Studies*, *8*(2), 426–439. https://doi.org/10.36892/ijlls.v8i2.2596

Sumardi, & Guci, R. I. (2023). HOTS-based language assessment literacy: Challenges and prospects in English language teaching. *Indonesian Journal of Applied Linguistics*, *12*(3), 818–827. https://doi.org/10.17509/ijal.v12i3.44261

Tian, Y., & Crossley, S. (2026). Linking writing processes to argument elements: Insights from keystroke logging in L1 and L2 writers. *Journal of Second Language Writing*, *71*. https://doi.org/10.1016/j.jslw.2025.101276

Tzoannopoulou, M. (2015). Rethinking ESP: Integrating content and language in the university classroom. *Procedia - Social And Behavioral Sciences*, *173*, 149–153. https://doi.org/10.1016/j.sbspro.2015.02.045

Valentine, K. D., & Kopcha, T. J. (2013). Conducting educational design research. In *Educational Media International* (Vol. 50, Number 3). https://doi.org/10.1080/09523987.2013.843832

Waluyo, B., Zahabi, A., & Ruangsung, L. (2024). Language assessment at a Thai university: A CEFR-based test of English proficiency development. *REFLections*, *31*(1).

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, *9*(1), 27–55. https://doi.org/10.1016/j.asw.2004.01.002

Ye, W., & Liu, J. (2023). Exploring and modeling the reading-writing connection in EFL integrated writing. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1161272