



EXPLORING ENGLISH PROFICIENCY LEVELS OF INDONESIAN UNIVERSITY STUDENTS THROUGH ENGLISHSCORE BASED ASSESSMENT

Diah Kurniati¹, Ahdi Riyono², Achmad Solaeman³, Arceli R. Millan⁴, Nooreen Noordin⁵

^{1,2,3}Universitas Muria Kudus, Indonesia

⁴Nueva Ecija University of Science and Technology, Philippines

⁵Universiti Putra Malaysia, Malaysia

Corresponding author: diah.kurniati@umk.ac.id

APA Citation: Kurniati, D., Riyono, A., Solaeman, A., Millan, A. R., & Noordin, N. (2025). Exploring English proficiency levels of Indonesian university students through EnglishScore based assessment. *Indonesian EFL Journal*, 11(3), 539-552. <https://doi.org/10.25134/ieflj.v11i3.54>

Received: 12-08-2025

Accepted: 02-10-2025

Published: 30-11-2025

Abstract: English language proficiency is a critical competency for university students in Indonesia, serving as a gateway to academic achievement, global engagement, and professional advancement. Despite its importance, limited empirical data exist on students' actual proficiency levels, especially through mobile-based assessments designed to reflect real-world language use in context. This study investigates the English proficiency of Indonesian undergraduate students using the EnglishScore mobile application—an internationally recognized assessment tool that differs from traditional exams like TOEFL or IELTS by offering short, adaptive, and skills-specific tests accessible via smartphones. Utilizing a quantitative descriptive design, the study assessed 1,216 students from various departments on grammar, vocabulary, reading, and listening, aligned with CEFR standards. Results show that 55% of students are at A1, 28% at A2, 12% at B1, and only 3% at B2. These findings reveal a substantial gap between expected and actual proficiency levels after years of formal instruction. The research highlights the urgent need for universities to adopt communicative, proficiency-oriented curricula and leverage mobile-based diagnostics to inform instruction. Mobile tools like EnglishScore, which mirror authentic language use by simulating everyday communicative tasks, offer scalable, low-barrier alternatives for assessment. This study contributes to the underexplored area of digital language testing in Indonesia and provides actionable insights for policymakers and educators to enhance language learning strategies through technology-informed policy and pedagogy.

Keywords: *English proficiency; university students; EnglishScore; language assessment; CEFR.*

INTRODUCTION

English competence is essential for academic progress and global engagement, prompting the worldwide implementation of the CEFR as a defined benchmark for quantifiable language results (Maral & Ergül, 2025; Kim et al., 2025). Its hierarchical descriptors help set expectations for grammar, vocabulary, reading, and listening at different levels, and it has wordlists and graded text difficulty based on a corpus (Wwolver & Lew, 2025). Indonesian learners still don't meet global standards, nevertheless. The EF EPI always puts Indonesia in the low-proficiency band (A2–B1), which shows that there are big gaps in basic language skills (Arisandi & Irawan, 2023;

Khushik, 2025). Recent reforms, like Kurikulum Merdeka, put a lot of emphasis on communicative competence, but they don't have a systematic way to connect with the CEFR or give national-level diagnostic evidence of how well students are doing (Cooper, 2025; Sanemueang, 2025). As a result, scalable mobile-based CEFR tests like EnglishScore can help us get standardized, population-level information about how well Indonesian university students speak English (Nadtayay & Wongsaphan, 2024).

Recent research on CEFR lexical profiling has transformed the definition of competency by anchoring level descriptors in empirical frequency, polysemy, and dictionary-view data (Soliman &

Familiar, 2024). Sentence-level CEFR prediction algorithms utilize syntactic and morphological metrics to classify learner texts with increasing precision (Uchida et al., 2024). Critically, linguistic complexity metrics, covering lexical sophistication and syntactic embedding, allow educators to connect instructional materials with CEFR thresholds in a more evidence-based manner (Zhang & Lu, 2025). This corpus-driven improvement strengthens the interpretative validity of CEFR assessment problems (Wolfer & Lew, 2025). Building on these achievements, generative AI adds lexical and structural information to further strengthen CEFR predictions for learner outputs (Uchida, 2025; Uchida & Negishi, 2025).

CEFR's influence on curricular design reflects a global movement for transparent standards in grammar, reading, and communication (Kim et al., 2025). Systematic reviews reveal that CEFR alignment increasingly guides textbook production, teacher training, and syllabus sequencing (Viora & Surya, 2025). Regionally, bibliometric research maps rapid CEFR adoption across ASEAN, showing a collective move toward internationally benchmarked English teaching (Nadtayay & Wongsaphan, 2024). Beyond Europe, CEFR adaption for Arabic, which integrates multi-dialectal frequency analysis, shows its versatility for multiple language situations (Pratama et al., 2024). These advances support CEFR as a widely acknowledged basis for defining proficiency advancement (Soliman & Familiar, 2024).

Digital CEFR-aligned learning platforms offer structured routes for vocabulary and comprehension development through gamified instruction and adaptive feedback (Sanemueang, 2025). Listening and reading difficulty prediction utilizing machine-learning increases the precision of CEFR-level classification, enabling more focused resource selection (Cooper, 2025). Frequency-augmented digital dictionaries complement CEFR vocabulary lists by optimizing level-consistent lexical exposure (Wolfer & Lew, 2025). Importantly, AI-supported adaptive systems operationalize CEFR as a dynamic framework rather than a static standard, allowing materials to modify according to learner trajectories (Uchida, 2025). This trend reveals how digital ecosystems progressively ground their teaching rationale in CEFR concepts.

Generative AI offers expanded prospects for CEFR-aligned automated scoring, delivering fresh validity evidence for large-scale formative

evaluation (Pack et al., 2024). Teacher educators accept its potential to expedite assessment, while issues exist about academic integrity and suitable pedagogical integration (Moorhouse & Kohnke, 2024). Learners generally evaluate AI writing support positively for grammar and structure, contingent on transparent usage instructions (Ho, 2024; Lee et al., 2024). Scholars note AI's democratizing promise for equal access to English learning, especially in under-resourced situations (Tafazoli, 2024). At the same time, concerns regarding bias and model opacity underline the necessity for rigorous validation before integrating AI into CEFR evaluation ecosystems (Crompton et al., 2024).

The development of online evaluation has enhanced CEFR's significance as the backbone of digital assessment frameworks. AI-assisted listening technologies greatly increase diagnostic accuracy, guaranteeing more consistent measurement across learners (Aryadoust et al., 2024). Frequent online assessment facilitates progressive monitoring and fosters student autonomy (Furwana et al., 2024). Pedagogical technologies, including multimedia scaffolding and organized reading modules, enhance CEFR-aligned competency when incorporated coherently (Djalilova & Kazonov, 2024; Cappellini, 2024). Strategic communicative competence frameworks further underline the usefulness of CEFR-based multimodal assessment approaches (Malykhin et al., 2024; Saban-Dülger & Özcebe, 2024). Together, these examples indicate a progressive convergence between CEFR and digital assessment infrastructures.

Reading skill, particularly at CEFR B2 and above, is a defining indicator of academic preparation in English-medium classrooms (Warnby, 2025). Empirical data reveals that poor CEFR-aligned reading abilities associated with lower academic accomplishment in higher education (Ruegg et al., 2024). AI-supported CEFR difficulty prediction supports instructors in picking texts that fit learners' developmental levels (Cooper, 2025). Vocabulary profiling study explains lexical thresholds necessary for academic comprehension (Wolfer & Lew, 2025). These findings underline the relevance of CEFR reading descriptors for university transfer (Uchida et al., 2024).

Listening development involves a complex combination of cognitive load, lexical access, and contextual familiarity (Chachar et al., 2024; Yang et al., 2024). Adaptive, AI-driven listening assessments reduce performance variance by

altering difficulty in accordance with CEFR bands (Aryadoust et al., 2024). Affective characteristics like as resilience and emotion management strongly influence outcomes (Zhi & Derakhshan, 2024). Sociolinguistic variety and exposure disparities significantly complicate listening performance, especially in multilingual learning environments (Nguyen et al., 2021; Misiran et al., 2018). These issues indicate to the necessity for CEFR-based listening evaluation methods that remain attentive to learner variety.

Motivational factors show a crucial influence on CEFR-based proficiency progress, with great academic buoyancy related with more consistent developmental gains (Yang et al., 2024). Vocabulary programs such as Duolingo provide independent, CEFR-referenced lexical reinforcement (Fakhrurriana et al., 2024). Creative writing assignments boost learners' syntactic control and lexical breadth in ways that map onto higher CEFR bands (Fitria, 2024). Likewise, AI-based speaking chatbots add to oral fluency and interaction competency (Yuan, 2024). The ability to regulate boredom and maintain engagement is similarly crucial for sustaining learning trajectories inside CEFR-guided online settings (Pawlak et al., 2025).

Sociocultural contexts impact learners' access to English exposure, with socioeconomic background influencing motivation and language trajectories (Yokubjonova, 2025). Authentic communication possibilities in tourism environments give functional input that increases competency (Ahmed et al., 2024; Mukharramkhon & Parizoda, 2024). Internet-mediated exposure enhances vocabulary and promotes independent learning in circumstances with less teaching (Curle et al., 2024). Analyses of English hegemony emphasize how global linguistic hierarchies influence student ambitions and educational policies (Zeng & Yang, 2024). These aspects jointly underline the requirement of CEFR assessment techniques capable of capturing proficiency across multiple sociocultural contexts (Hossain, 2024).

English competence highly predicts academic performance in university programs requiring substantial reading, writing, and critical engagement (Ruegg et al., 2024). Indonesian graduates typically indicate that low proficiency hampers employability and access to global mobility prospects (Taufik, 2024). English-medium instruction research demonstrate that insufficiency in CEFR-aligned abilities increases cognitive burden and impedes understanding

(Curle et al., 2024). Indonesian students generally attend tertiary education at A2–B1 levels, reducing their academic engagement (Erzad & Suciati, 2024). Persistent instructional and affective hurdles further increase stagnation (Devi, 2023; Renandya et al., 2018).

National studies continuously demonstrate systemic deficiencies in grammar accuracy, vocabulary breadth, and receptive abilities, with most Indonesian university students clustering around A2–B1 equivalents (Damayanti et al., 2020; Prawiyata, 2020). Textbook studies suggest poor CEFR alignment, limiting learners' access to graded input (Rais & Awwalia, 2025). Broader reviews indicate structural constraints like limited exposure, uneven pedagogy, and lack of communicative practice (Rofik et al., 2020). Affective issues, such as worry and low confidence, further hinder growth (Zarate, 2022; Misiran et al., 2018). Collectively, these data highlight an urgent need for scalable CEFR diagnostic systems capable of exposing true proficiency distributions.

Although substantial research investigates CEFR structures, AI-supported learning settings, and digital assessment technologies, no Indonesian study has systematically analyzed CEFR proficiency distributions using a mobile-proctored, AI-assisted test such as EnglishScore. Existing studies depend mostly on TOEFL or classroom-based examinations that lack CEFR compatibility, do not measure vocabulary, grammar, reading, and listening in an integrated framework, and are not scalable for large cohorts. Consequently, a crucial gap persists in getting authentic, population-scale CEFR-aligned proficiency data that reflects genuine learner performance.

This study does not analyze speaking, classroom pedagogy, or cross-test comparisons, concentrating only on CEFR-level outcomes from EnglishScore's grammar, vocabulary, reading, and listening modules. This research is the first in Indonesia to apply EnglishScore, a mobile, AI-enhanced CEFR evaluation incorporating adaptive difficulty, digital proctoring, and automated analytics, to map university students' competency distributions. This methodological innovation enhances large-scale CEFR diagnostic practice beyond traditional paper-based or lab-based testing.

The study presents empirical CEFR-level evidence needed for curriculum development, placement decisions, and national policy planning. By establishing the possibility of mobile CEFR diagnostics for large cohorts, it promotes

Indonesia's move toward transparent, egalitarian, and internationally compatible English proficiency evaluation.

METHOD

This study employed a descriptive quantitative research design aimed at providing a comprehensive overview of university students' English language proficiency. Descriptive quantitative research allows for the systematic collection and numerical analysis of data to identify patterns, trends, and distributions within a defined population (Creswell & Creswell, 2017). Through this methodological approach, the study sought to quantify language proficiency levels and present objective, data-driven insights that are critical for educational assessment and policy formulation.

The research was conducted at a private university located in the Kudus Regency of Central Java Province, Indonesia. The participants consisted of 1,216 second- and third-year undergraduate students enrolled in a compulsory English course. These students represented a diverse array of academic disciplines, including Mechanical Engineering, Agricultural Technology, Guidance and Counseling, Law, Management, Indonesian Language and Literature Education, English Language Education, Primary School Teacher Education, Management, Accounting, Psychology, Mathematics Education, Information System, Electrical Engineering, Informatics Engineering. The distribution of the students can be seen in the following figure.

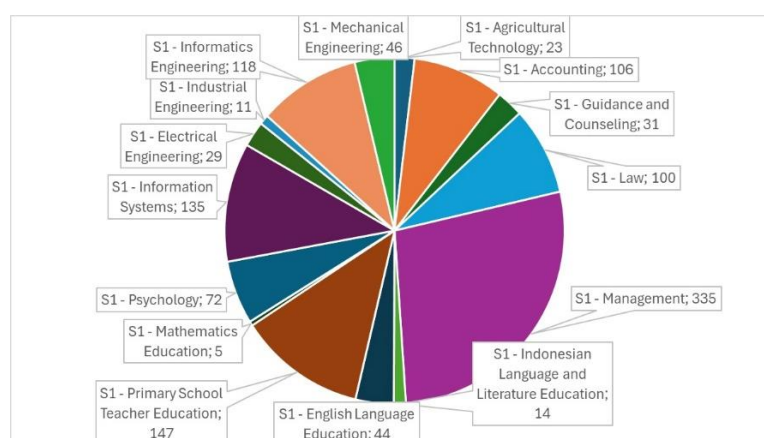


Figure 1. *Students' distribution at each study program*

This variety of study programs shows that participants' English proficiency levels may have varied, impacted by the specific language needs and exposure within their different disciplines. Prior to taking the EnglishScore test, all participants completed 16 instructional sessions aiming at providing a common foundation and improving their overall English language skills, so helping reduce disparities in language competence among them.

The participants were distributed across 27 different classes, taught by a total of 10 lecturers. This varied academic background provided a broad and representative sample for evaluating English proficiency levels across different fields of study.

The primary data collection instrument in this study was the EnglishScore Test, a mobile-based language proficiency assessment developed by the British Council. Specific requirements are imposed for each EnglishScore exam taken on a mobile

device in order to avoid cheating and to foster reliability and trust. As part of this, test takers must log in and consent to exam conditions, which include being photographed during the session. To further preserve the integrity of the exam, a variety of behaviors are also forbidden throughout the assessment. EnglishScore is the world's only mobile-proctored English language test. Proctoring is done by taking photos at random intervals during each test session and securely keep them in the cloud. (See Figure 2). These photographs are then included in a security report with the professional certificate, demonstrating that the test-taker is alone and unassisted during the test. The presence of an accessible camera during the assessment reduces fraud and cheating. (<https://www.englishscore.com/wp-content/uploads/2022/05/Security-Report-2021.pdf>)



Figure 2. *Proctoring photos captured during a sitting of the EnglishScore test*

As it was mentioned earlier that EnglishScore uses real-time facial recognition and camera monitoring technology to detect and respond to potentially unfair exam behavior. The system automatically pauses the test in the following conditions (See Figure 3):

Face Not Detected. If the test-taker's face is not clearly visible in the camera view (for example, because they have turned away), the test is paused with a warning.

Multiple Faces Detected. If the system detects more than one person in the camera frame, the test is delayed to avoid possible collaboration or impersonation.

Not Looking at the Screen. If the test-taker repeatedly looks away from the screen, the test is halted to maintain focus and discourage the use of illegal resources.

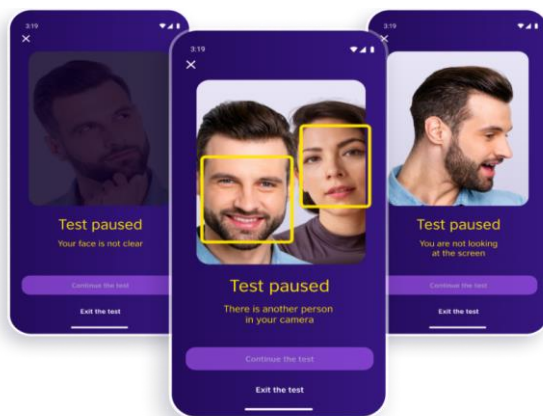


Figure 3. *EnglishScore AI-Powered test monitoring for secure online assessment*

Englishscore tests have also demonstrated their validity and reliability. Validity refers to whether a test accurately measures what it is designed to assess, and EnglishScore was created with a significant focus on psychometric validity. This comprises concept validity, criterion-related validity, and face validity, all of which are required for a reliable and equitable language evaluation. For construct validity, EnglishScore assesses general English ability in grammar, vocabulary, reading, and listening, utilizing adaptive testing technology that adapts question difficulty based on test-taker responses. This guarantees that the exam

targets the right difficulty level for each user and is consistent with known frameworks such as the CEFR (Common European Framework of Reference).

In terms of criterion-related validity, the British Council's research demonstrate that EnglishScore scores substantially correlate with other standardized English exams such as IELTS and TOEFL, with correlation values ranging from 0.7 to 0.85 (British Council, 2020). These data verify the test's consistency with widely accepted proficiency measures. EnglishScore also displays great face and content validity by including actual academic and workplace circumstances in its tests, which increases relevance and user engagement. All items are rigorously developed and tested by language assessment professionals to guarantee they are psychometrically sound and appropriately measure the target language constructs.

Reliability is defined as the consistency and stability of test scores across time or across test versions. The EnglishScore platform is highly reliable, as evidenced by its internal consistency and test-retest dependability. Internally, EnglishScore uses a computer-adaptive testing engine that adjusts the level of difficulty of questions depending on the test-taker's answers. This adaptive technique improves measurement accuracy while minimizing random errors, yielding high internal consistency. Reported reliability coefficients (Cronbach's alpha) are often more than 0.85, indicating high internal reliability (British Council 2020). In terms of test-retest reliability, research have shown that when test-takers repeat the test, their scores tend to be similar, as long as their proficiency does not change much. Although specific questions change due to the test's adaptive nature, the overall estimate of language competence remains consistent (British Council, 2020).

The EnglishScore test used in this study focused on core language skills—namely grammar, vocabulary, reading, and listening is administered at the end of the course. These components served as key indicators of the students' English language competencies. In preparation for the assessment, participants were instructed to download the

EnglishScore mobile application from the Google Play Store (Android) or the Apple App Store (iOS), create an account, and prepare for the test. The test duration ranged from approximately 30 to 40 minutes, and upon completion, students received immediate feedback in the form of a numerical score. These scores were used to evaluate and classify students' proficiency levels in accordance with standardized metrics (Bachman & Palmer, 1996).

The quantitative data obtained from the EnglishScore Test were analyzed using descriptive statistical methods. Measures such as the mean, standard deviation, frequency distribution, and percentage were employed to summarize and interpret the test results. This analytical approach enabled a clear depiction of the students' performance across various proficiency levels. The results were presented through a combination of tables and graphical illustrations to enhance readability and facilitate a holistic understanding of the findings (Field, 2024). Such visualization also supports transparent communication of trends and comparative analysis among different student groups.

This study adhered to rigorous ethical research standards to ensure the integrity and confidentiality

of participant data. Although participation in the assessment was integrated into a mandatory course component, all test scores and personal identifiers were anonymized to protect student privacy. No identifying information was disclosed in the reporting of results. The data were utilized solely for academic and research purposes, with the aim of enhancing pedagogical strategies and contributing to the broader understanding of English language proficiency among university students (Abbas et al., 2016). Ethical clearance was aligned with institutional guidelines governing student research participation and data protection.

RESULTS AND DISCUSSION

Indonesian university students' language proficiency levels

The results of English proficiency test performed by using EnglishScore provide useful information about the English language proficiency of Indonesian university students. The proficiency levels of students are classified according to the Common European Framework of Reference for Languages (CEFR), which ranges from A1 (beginner) to B2 (Independent user). The details can be seen in Table 1.

Table 1. Students' language proficiency measured through EnglishScore

CERF	Numbers	Average	Minimum	Maximum	Std Dev
A1	666	149,64	104	199	26,72
A2	337	249,02	200	297	30,36
B1	174	333,78	300	390	24
B2	39	433,97	400	494	25,4

Meanwhile the distribution percentage of the score can be seen in the following figure

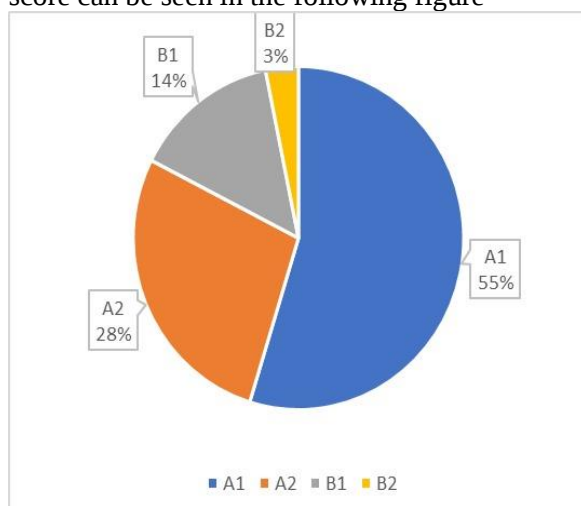


Figure 4. The distribution percentage of English proficiency level

Table 1 and Figure 4 illustrate that a significant proportion of students exhibiting limited English proficiency are categorized within the A1 and A2 levels (Basic User), with only a minor fraction attaining B1 and B2 levels (Independent User), which are generally deemed essential for academic and professional achievement. The lack of C1/C2 level outcomes indicates that no students fulfilled the criteria for the proficient user classification.

The current investigation, encompassing a sample of 1.216 students, identified that 666 students (55%) were situated at the A1 level, with scores ranging from 104 to 199 and an average score of 149.64. This finding underscores fundamental English proficiency, characterized by rudimentary vocabulary and grammatical structures, thereby enabling students to comprehend and utilize basic phrases. In contrast, 337 students (28%) were classified at the A2 level, with scores varying from 200 to 297 (average:

249.02), reflecting advancement yet still encountering difficulties with complex grammar and communication. A total of 174 students (14%) were designated as B1 level, with scores spanning from 300 to 390 (average: 333.78), signifying the capacity to engage in intermediate English communication. Lastly, only 39 students (3%) achieved the B2 level, with an average score of 433.97, illustrating the ability to interact fluently with native speakers, grasp fundamental concepts in sophisticated texts, and produce extensive written work.

The findings of the present investigation align with antecedent studies indicating that the English language proficiency of Indonesian university students is situated within the basic user or beginner classification (Damayanti et al., 2020; Prawiyata, 2020; Rais & Awwalia, 2025; Rofik et al., 2020). In their study, (Damayanti et al., 2020) employed the Test of English for International Communication (TOEIC) to assess the English competencies of university scholars. Their analysis revealed that a significant proportion of Poltekba students attained scores below 400, which corresponds to an A2 level on the Common European Framework of Reference (CEFR). Furthermore, (Prawiyata, 2020) determined that the TOEFL scores of postgraduate students were categorized within the Low Intermediate Level. The current research, which identified basic user proficiency among university students at a private institution in Central Java province, Indonesia, corroborates the findings presented in the study conducted by (Rais & Awwalia, 2025) regarding university students' proficiency in Sintang, West Kalimantan. (Rais & Awwalia, 2025) utilized a series of distributed multiple-choice questions from the British Council English framework to evaluate the language proficiency of students. Moreover, parallels can be drawn between the outcomes of this study and the English proficiency levels of Thai university students, as documented by (Nithideechaiwarachok et al., 2022), who similarly reported that their participants encountered difficulties with various aspects of the English language, including vocabulary, grammar, pronunciation, speaking, and writing. Both cohorts were assessed to have English proficiency levels classified as A2, or basic user.

Nevertheless, linguistic competence in various non-native English-speaking nations exhibited divergent results from this investigation. As per the 2024 EF English Proficiency Index, Singapore attained a commendable third position worldwide with a score of 609, which signifies a category of

very high proficiency. The Philippines closely trailed at the 22nd position, achieving a score of 570, which reflects a high level of proficiency. Malaysia also recorded a score of 566, placing it 26th globally and categorizing it within the high proficiency band. Conversely, Indonesia was positioned 80th with a score of 468, thus categorizing it within the low proficiency spectrum. This analysis indicates that Singapore, the Philippines, and Malaysia exhibit notable English proficiency, whereas Indonesia demonstrates a comparatively lower score within the regional context.

Researchers have elucidated various determinants that influence English proficiency among university scholars. (Misiran et al., 2018) delineated linguistic, cultural, and educational variables as pivotal factors affecting the English proficiency levels of Malaysian students. In the Philippines, (Tafazoli, 2024) ascertained those intrinsic elements such as anxiety, attitude, motivation, and multiple intelligences, alongside extrinsic factors including teaching efficacy, media exposure, and pedagogical methodologies, exert a significant influence on students' English proficiency. Concurrently, (Ly., 2024) revealed that the inadequate English competency among Indonesian EFL learners was attributable to both internal factors (namely low motivation and learning anxiety) and external determinants (such as inappropriate teaching strategies, large classroom sizes, and unfavourable English learning environments). Malykhin et al. (2024) also identified six internal factors (namely English Knowledge, Motivation, Learning Preferences, Fear of Speaking English, English Language Experience, and Perception of English) and four external factors (including the Role of Technology and Media, Classroom Method Preferences, Environmental Factors, and the Complexity and Difficulty of English) that affect English proficiency among university students in Indonesia. In a similar vein, (Nguyen et al., 2021) underscored the significance of intrinsic factors such as learning objectives, abilities, methodologies, perseverance, and self-awareness, as well as extrinsic factors such as demographic variables, parental influence, and the educational milieu, in shaping the English proficiency of Vietnamese students. Collectively, these investigations illuminate the intricate challenges associated with English language acquisition within higher education across diverse cultural and educational frameworks.

Score variability

In accordance with score variability, the results of this study show that the A2 group have the highest standard deviation (30.36), indicating more variety in scores. The B1 and B2 groups have lower standard deviations (24.00 and 25.43), indicating greater level proficiency. The A1 group shows moderate variation (26.72), indicating that there are noticeable language proficiency differences even among beginners.

The study results can be explained in detail based on the language skills and component, namely, grammar, vocabulary, reading, and listening.

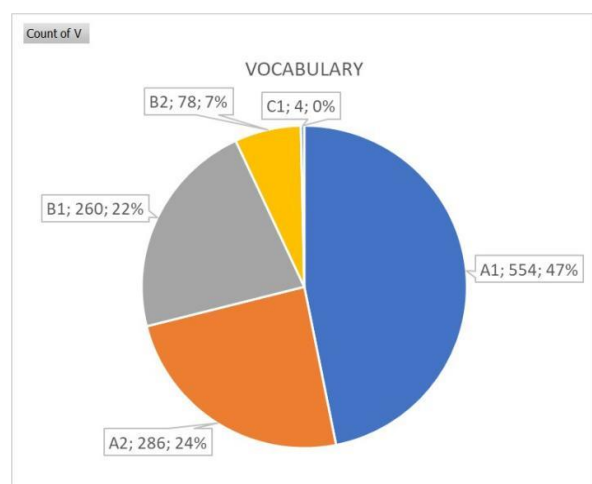


Figure 5. *Students' English vocabulary level*

The figure presented below illustrates the proficiency levels of students' vocabulary based on the EnglishScore assessment. Almost half of the cohort, specifically 47% or 554 individuals, are classified at the A1 level, which signifies that they possess a limited lexicon and comprehend only fundamental words and phrases. Concurrently, 24%, or 286 individuals, attained the A2 level, indicating that they are familiar with frequently utilized terms but face difficulties with less common or academic vocabulary. Moreover, 22%, or 260 individuals, achieved a score corresponding to the B1 level, suggesting that they possess a more varied vocabulary, although they may still encounter challenges with less explicit meanings. Nevertheless, merely 7%, or 78 individuals, reached the B2 level, reflecting that only a small fraction demonstrates proficient vocabulary knowledge requisite for advanced reading and writing tasks. Ultimately, the C1 (Advanced) level is virtually absent, with only four individuals attaining this level of proficiency, thereby indicating that the vast majority of students exhibit a deficiency in advanced vocabulary skills.

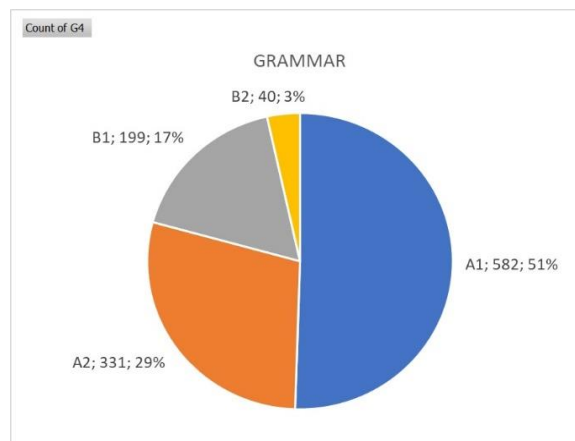


Figure 6. *Students' grammar level*

Figure 6 illustrates students' levels of grammatical proficiency as measured by the EnglishScore test. The majority of students, 51% (582 students), are at A1 level, suggesting a basic knowledge of English grammar with frequent errors in sentence structure, verb tenses, and word order. Meanwhile, 29% (331 students) reached the A2 level, meaning they have some grasp of basic grammatical structures, yet they still struggle with more complex sentences. Furthermore, 17%, or 199 students, scored at the B1 level, showing that they can construct more accurate sentences, although they may still make errors with conditionals, modals, and complex tenses. However, only 3%, or 40 students, achieved the B2 level, indicating that few can utilize grammar successfully in both written and spoken communication.

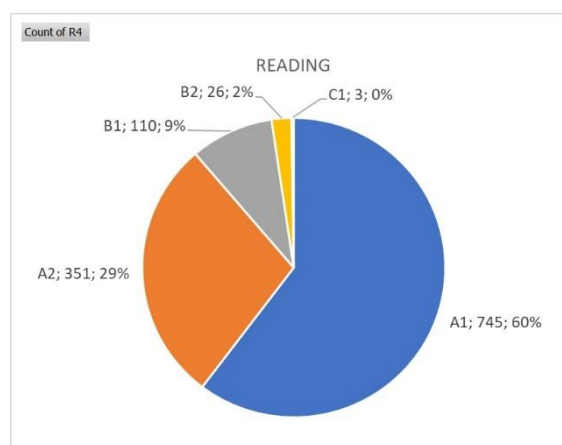


Figure 7. *Students' reading level based on CEFR*

Figure 7 explains the students' English reading level. The majority of students, 60% (745 students), are at A1 level, which indicates basic reading skills with limited vocabulary and understanding of simple texts. Meanwhile, a significant portion, 29% or 351 students, achieve A2 level, indicating that they can understand short, simple texts but struggle with complicated

sentences and academic reading. Furthermore, a smaller proportion, 9% or 110 students, obtain B1, indicating some capacity to understand longer texts and extract key elements, while only 2% or 26 students get B2, indicating the ability to manage more complicated texts and infer meaning. Finally, only three students have achieved C1.

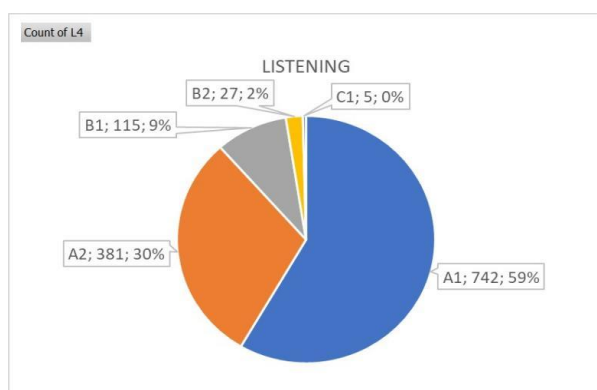


Figure 8. Students' listening level based on CEFR

Based on the results shown in Figure 8, the majority of students, 59% (742 students), are at the A1 level, which means they can only understand very simple words and phrases spoken slowly. Meanwhile, 30%, or 381 students, achieved the A2 level, meaning that they can grasp familiar expressions and phrases, but they still struggle with longer conversations. Additionally, 9%, or 115 students, are at the B1 level, suggesting they can understand general speech but may struggle with accents and fast-paced conversations. However, just 2%, or 27 students, achieved B2, indicating that few can follow complicated discussions or academic lectures. The pie graphic depicts students' listening proficiency levels. Finally, just 5 students achieve C1 (Advanced) proficiency, indicating that nearly all students struggle with high-level listening comprehension.

Distribution of proficiency levels across skills
Students' proficiency levels in core language skills can be seen in details in Table 2

Table 2. Distribution of proficiency levels across skills

Proficiency Levels	Grammar	Vocabulary	Listening	Reading
A1	51%	47%	59%	60%
A2	29%	24%	30%	29%
B1	17%	22%	9%	9%
B2	3%	7%	2%	2%
C1	0%	0%	0%	0%

According to the table above, Reading and Listening are the most difficult skills for students with 89% of students remaining at the A1–A2 levels followed by Grammar with 80% of students also remains at the A1-A2 levels. Meanwhile, vocabulary with the lowest percentage (71%) at A1–A2, indicating relatively stronger proficiency. Students struggle with reading and listening comprehension for a variety of reasons. These include issues with spoken and written language speed and length, a lack of prior knowledge, distractions, and psychological components like anxiety and insecurity.

In listening test, students may struggle with the speed at which information is given, especially in spoken language, making it challenging to comprehend and remember knowledge. Longer spoken or written materials can cause weariness and loss of concentration. Similar to this, in reading test, students may become overwhelmed by lengthy texts or sections with intricate sentence structures, making it difficult for them to concentrate and comprehend the main idea.

Moreover, when a student is unfamiliar with the topic or context of the hearing or reading material, understanding suffers greatly. Background knowledge provides a context for understanding incoming information. Listening and reading comprehension can also be adversely affected by anxiety, a fear of making mistakes, and a lack of confidence. These elements may contribute to elevated stress levels, diminish concentration and hinder reading comprehension. Finally, students may find it challenging to focus on the listening material due to distractions from outside noise, personal electronics, or even their own thoughts.

CONCLUSION

This study assessed the English proficiency of Indonesian university students using the EnglishScore mobile app, revealing a concerning trend. The majority of students were classified at the A1 (Beginner) and A2 (Elementary) levels of the CEFR, with only a small minority reaching the independent B1 or B2 levels. This significant gap between expected and actual competence after years of formal education underscores a systemic

issue in the language curriculum from primary through tertiary levels.

The findings urge a shift in English language education. Educators should adopt more communicative, proficiency-based approaches, moving away from exam-centric methods. For students, this highlights the need to take greater ownership of their learning by leveraging digital tools for self-assessment and practice. Institutions and policymakers are implored to redesign curricula and integrate mobile-based assessments to better align with international standards.

A key limitation was the focus only on receptive skills (reading, listening, grammar, and vocabulary), excluding productive skills like speaking and writing. Future efforts should pilot the integration of tools like EnglishScore into institutional assessments and conduct longitudinal studies to track progress. Adopting these innovations is crucial for equipping Indonesian students with the English skills needed for global success.

REFERENCES

- Abbas, A., Brennan, J., Abbas, J., Gantogtokh, O., & Bryman, K. (2016). *Teaching excellence in the disciplines*. The London School of Economics and Political Science.
- Ahmed, A. R., Abdulhamid, A. R., Khaled, A. R., & Wagdi, B. H. (2024). The importance of English language in tourism sector: A study in Socotra Island. *Journal of Language Learning and Assessment*, 85–90. <https://doi.org/10.71194/jlla.v2i2.116>
- Aryadoust, V., Zakaria, A., & Jia, Y. (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence*, 6(1), 1-10. <https://doi.org/10.1016/j.caeai.2024.100204>
- Bachman, L. F. (2024). Ongoing challenges in language assessment 1.1. In *The writings of Lyle F. Bachman* (pp. 438–458). Routledge.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bekou, A., Mhamed, M. B., & Assissou, K. (2024). Exploring opportunities and challenges of using ChatGPT in English language teaching (ELT) in Morocco. *Focus on ELT Journal*, 6(1), 87–106. <https://doi.org/10.1002/9781118411360.wbcla128>
- Cappellini, M. (2024). *Balancing reading and language learning: A resource for teaching English language learners, K-5*. Routledge.
- Cooper, C. R. (2025). Predicting the CEFR level of English listening texts with machine learning methods. *Research Methods in Applied Linguistics*, 4(3), 1-10. <https://doi.org/10.2307/328718>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503–2529. <https://doi.org/10.14744/felt.6.1.7>
- Damayanti, L., Gafur, A., & Balikpapan, P. N. (2020). English proficiency of students at Politeknik Negeri Balikpapan Based on TOEIC. *Teknosastik*, 18(1), 2020-2030. <https://doi.org/10.64152/10125/66620>
- de Winter, J. C. (2024). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 34(3), 915–930. <https://dx.doi.org/10.2139/ssrn.5202531>
- Dikaprio, V., & Diem, C. D. (2024). How effective is Talkpal. ai in enhancing English proficiency? Insights from an experimental study. *Language, Technology, and Social Media*, 2(1), 48–59. <https://doi.org/10.1111/bjet.13460>
- Djalilova, S., & Kazoqov, R. (2024). Enhancing English language learning through pedagogical technologies. *Modern Science and Research*, 3(2), 130–140. <https://doi.org/10.33365/ts.v18i1.452>
- Douglas, D. (2025). ESP and assessment. In *The handbook of English for specific purposes* (pp. 185–204). Sage publications.
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6(1), 1-9.
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6(1), 1-10. <https://doi.org/10.1016/j.caeai.2024.100230>
- Fakhrurriana, R., Nisa, A., & Noni, N. (2024). The perceptions of using Duolingo application in learning English for student's vocabulary mastery. *BATARA DIDI: English Language Journal*, 3(1), 45–54. <https://doi.org/10.56209/badi.v3i1.82>
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
- Fitria, T. N. (2024). Creative writing skills in English: Developing student's potential and creativity. *EBONY: Journal of English Language Teaching, Linguistics, and Literature*, 4(1), 1–17. <https://doi.org/10.37304/ebony.v4i1.10908>
- Furwana, D., Muin, F. R., Zainuddin, A. A., & Mulyani, A. G. (2024). Unlocking the potential: Exploring

- the impact of online assessment in English language teaching. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 12(1), 653–662. <https://doi.org/10.24256/ideas.v13i1.6765>
- Hasanah, U., & Wahyuni, S. (2025). Content standard analysis in the formative assessment of textbook English for Nusantara. In *Bookchapter English Language Teaching, Literature, and Translation*, 1(1), 458–491. <https://doi.org/10.15294/eltlt.v1i1.439>
- Ho, P. X. P. (2024). Using ChatGPT in English language learning: A study on IT students' attitudes, habits, and perceptions. *International Journal of TESOL & Education*, 4(1), 55–68. <https://doi.org/10.54855/ijte.24414>
- Hossain, K. I. (2024). Reviewing the role of culture in English language learning: Challenges and opportunities for educators. *Social Sciences & Humanities Open*, 9(1), 1–7. <https://doi.org/10.1016/j.ssaho.2023.100781>
- Khushik, G. A. (2025). Examining the effects of alternate prompts on syntactic complexity in EFL learners' essays at CEFR levels A1, A2, and B1. *Language Assessment Quarterly*, 22(3), 283–305. <https://doi.org/10.1080/15434303.2025.2528019>
- Kiczowski, M., & Lowe, R. J. (2024). Native-speakerism in English language teaching: 'Native speakers' more likely to be invited as conference plenary speakers. *Journal of Multilingual and Multicultural Development*, 45(5), 1408–1423. <https://doi.org/10.1080/01434632.2021.1974464>
- Kim, J., Isbell, D. R., & Kim, K. (2025). The influence of CEFR in South Korea's 2022 national English curriculum: A focus on grammar. *English Today*, 1–6. <https://doi.org/10.1017/S0266078425000124>
- Lee, Y. J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies*, 14(1), 1–7. <https://doi.org/10.30935/ojcm/14195>
- Ly, C. K. (2024). Teachers' roles on English language teaching for promoting learner-centered language learning: A theoretical review. *International Journal of TESOL & Education*, 4(2), 78–98. <https://doi.org/10.54855/ijte.24425>
- Malykhin, O., Bondarchuk, J., Tersina, I., & Voitanik, I. (2024). Unlocking success: Strategic approaches to enhancing communicative competence in English learning. *Amazonia Investiga*, 13(76), 90–102. <https://doi.org/10.34069/AI/2024.76.04.8>
- Maral, B. N., & Ergül, H. (2025). CEFR proficiency levels of secondary school English textbooks' word lists in Türkiye: Secondary school English textbooks' word lists. *International Journal of Curriculum and Instruction*, 17(3), 892–909.
- Misiran, M., Mahmuddin, M., Yusof, Z. M., Affan Jaafar, I., Firdhaus Joferi, A., & Manap, S. (2018). *Exploring factors that affect english proficiency level among university students: A case study in Universiti Utara Malaysia*. Penerbit Akademia Baru.
- Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System*, 122, 1–7. <https://doi.org/10.1016/j.system.2024.103290>
- Mukharramkhon, N., & Parizoda, M. (2024). Tourism as a tool for English language acquisition. *QO 'Qon Universiteti Xabarnomasi*, 13, 120–122. <https://doi.org/10.54613/ku.v13i.1038>
- Nadtayay, N., & Wongsaphan, M. (2024). Bibliometric analysis using VOSviewer with Publish or Perish of CEFR-based comparison of English language teaching models for communication. *ASEAN Journal of Educational Research and Technology*, 4(1), 1–10.
- Nguyen, L., Nguye, H., Tran, T. T. N., Le, M. Q., & Ziyi, Z. (2021). Factors influencing English proficiency: An empirical study of Vietnamese students. *Management Science Letters*, 1779–1786. <https://doi.org/10.5267/j.msl.2021.2.004>
- Nithidechaiwarachok, B., Maneekanon, O., & Bubphada, T. (2022). Exploring english language proficiency, english language problems, and english needs among first year undergraduate students. *International Journal of Learning, Teaching and Educational Research*, 21(12), 272–289. <https://doi.org/10.26803/ijlter.21.12.15>
- Novawan, A., Walker, S. A., & Ikeda, O. (2024). The new face of technology-enhanced language learning (TELL) with artificial intelligence (AI): Teacher perspectives, practices, and challenges. *Journal of English in Academic and Professional Communication*, 10(1), 1–18. <https://doi.org/10.25047/jeapco.v10i1.4565>
- Numonova, M. (2024). Visual aids in enhancing the learning process in English language in a private school (Akhmad Education). *Qo 'Qon Universiteti Xabarnomasi*, 13, 309–312. <https://doi.org/10.54613/ku.v13i.1086>
- Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence*, 6, 1–8. <https://doi.org/10.1016/j.caeai.2024.100234>
- Pawlak, M., Derakhshan, A., Mehdizadeh, M., & Kruk, M. (2025). Boredom in online English language classes: Mediating variables and coping strategies. *Language Teaching Research*, 29(2),

- 509–534.
<https://doi.org/10.1177/13621688211064944>
- Pratama, A. W., Mahliatussikhah, H., Murtadho, N., & Kholisin, K. (2024). Opportunities and challenges of Common European Framework of Reference for Language (CEFR) standardization in Arabic language learning. *Arabi: Journal of Arabic Studies*, 9(1), 33–42.
<https://doi.org/10.24865/ajas.v9i1.774>
- Prawiyata, Y. D. (2020). The use of anagrams to increase students' english vocabulary mastery. *Jurnal Penelitian Pendidikan Bahasa Dan Sastra*, 5(1), 22–24.
<https://doi.org/10.32696/ojs.v5i1.397>
- Rais, B., & Awwalia, L. (2025). What is the student's level? Analyzing students' english proficiency levels in high education level. *SALEE: Study of Applied Linguistics and English Education*, 6(1), 1–10. <https://doi.org/10.35961/salee.v6i1.1762>
- Rofik, A., Christina, C., & Hidayah, A. (2020). An english proficiency study on students of developing islamic higher education. *Linguists: Journal Of Linguistics and Language Teaching*, 6(2), 86–96.
<https://doi.org/10.29300/ling.v6i2.3008>
- Ruegg, R., Hoang, H., & Petersen, N. (2024). The influence of English language proficiency test scores on the academic success of ESL undergraduate students. *Educational Research and Evaluation*, 29(3-4), 111–129.
<https://doi.org/10.1080/13803611.2024.2314533>
- Saban-Dülger, N. S., & Özcebe, E. (2024). Language assessment in bilingual Turkish-speaking preschoolers with developmental language disorders: A tutorial. *Perspectives of the ASHA Special Interest Groups*, 9(3), 655–673.
https://doi.org/10.1044/2024_PERSP-23-00067
- Sadiqzade, Z. (2025). Idiomatic expressions and their impact on lexical competence. *Journal of Azerbaijan Language and Education Studies*, 2(1), 26–33.
<https://doi.org/10.69760/jales.2025001002>
- Sanemueang, C. (2025). The effectiveness of integrating CEFR-aligned English instructional program with gamification. *Journal of Education and Learning*, 14(5), 123–134.
<https://doi.org/10.5539/jel.v14n5p123>
- Saptiany, S. G., & Prabowo, B. A. (2024). Speaking proficiency among English specific purpose students: A literature review on assessment and pedagogical approaches. *LITERACY: International Scientific Journals of Social, Education, Humanities*, 3(1), 36–48.
<https://doi.org/10.56910/literacy.v3i1.1392>
- Saputra, N., Antoni, R., Widodo, A., Solissa, E. M., & Arief, I. (2024). Improving foreign language proficiency in society by decision tree classification. In *AIP Conference Proceedings* (Vol. 3001, No. 1, p. 110005). AIP Publishing LLC.
- Shawaqfeh, A. T., Jameel, A. S., Al-adwan, L. A. Y., & Khasawneh, M. A. S. (2024). Interaction as a mechanism to enhance English language proficiency in the classroom. *Journal of Language Teaching and Research*, 15(1), 229–234. <https://doi.org/10.17507/jltr.1501.25>
- Soliman, R., & Familiar, L. (2024). Creating a CEFR Arabic vocabulary profile: A frequency-based multi-dialectal approach. *Critical Multilingualism Studies*, 11(1), 266–286.
- Susanto, A., Ratnasari, S. L., Susanti, E. N., Megah, S. I., Wilany, E., & Yuliani, S. (2024). Beliefs of English language instruction by Indonesian elementary school teachers: Exploring the influence of environment and educational background. *AL-ISHLAH: Jurnal Pendidikan*, 16(1), 1–13.
<https://doi.org/10.35445/alishlah.v16i1.4434>
- Tafazoli, D. (2024). Exploring the potential of generative AI in democratizing English language education. *Computers and Education: Artificial Intelligence*, 7, 1–8.
<https://doi.org/10.1016/j.caeai.2024.100275>
- Taufik, M. (2024). English language proficiency and career opportunities: Perceptions of Indonesian university graduates. *Language Value*, 17(1), 85–107. <https://doi.org/10.6035/languagev.7933>
- Twersky, S. E., Jefferson, R., Garcia-Ortiz, L., Williams, E., & Pina, C. (2024). The impact of limited English proficiency on healthcare access and outcomes in the US: A scoping review. *Healthcare*, 12(3), 364–374.
<https://doi.org/10.3390/healthcare12030364>
- Uchida, S. (2025). Generative AI and CEFR levels: Evaluating the accuracy of text generation with ChatGPT-4o through textual features. *Vocabulary Learning and Instruction*, 14(1), 2078–2078.
<https://doi.org/10.29140/vli.v14n1.2078>
- Uchida, S., & Negishi, M. (2025). Assigning CEFR-J levels to English learners' writing: An approach using lexical metrics and generative AI. *Research Methods in Applied Linguistics*, 4(2), 1–9. <https://doi.org/10.1016/j.rmal.2025.100199>
- Uchida, S., Arase, Y., & Kajiwarra, T. (2024). Profiling English sentences based on CEFR levels. *ITL-International Journal of Applied Linguistics*, 175(1), 103–126.
<https://doi.org/10.1075/itl.22018.uch>
- Viora, D., & Surya, Y. F. (2025). Common European Framework of Reference (CEFR): A systematic literature review and bibliometric. *Innovative: Journal of Social Science Research*, 5(4), 3150–3165.
<https://doi.org/10.31004/innovative.v5i4.20653>
- Warnby, M. (2025). Relating academic reading with academic vocabulary and general English proficiency to assess standards of students'

- university-preparedness—the case of IELTS and CEFR B2. *Scandinavian Journal of Educational Research*, 69(3), 506–523. <https://doi.org/10.1080/00313831.2024.2318434>
- Wolfer, S., & Lew, R. (2025). Supplementing CEFR-graded vocabulary lists for language learners by leveraging information on dictionary views, corpus frequency, part-of-speech, and polysemy. *Humanities and Social Sciences Communications*, 12(1), 1–11. <https://doi.org/10.1057/s41599-025-05446-y>
- Yang, S., Azari Noughabi, M., & Jahedizadeh, S. (2024). Modelling the contribution of English language learners' academic buoyancy and self-efficacy to L2 grit: Evidence from Iran and China. *Journal of Multilingual and Multicultural Development*, 45(7), 2701–2717. <https://doi.org/10.1080/01434632.2022.2062368>
- Yokubjonova, D. (2025). Exploring the impact of social backgrounds on English language learning: A study of Tadjik and Kyrgyz students. *International Journal of Artificial Intelligence*, 1(2), 705–712.
- Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments*, 32(10), 6774–6789. <https://doi.org/10.1080/10494820.2023.2282112>
- Yuldashova, N. (2024). English proficiency for advancing agricultural development. *Scientific and Technical Journal "Sustainable Agriculture"*, 22(2), 98–100.
- Zeng, J., & Yang, J. (2024). English language hegemony: Retrospect and prospect. *Humanities and Social Sciences Communications*, 11(1), 1–9. <https://doi.org/10.1057/s41599-024-02821-z>
- Zhang, X., & Lu, X. (2025). Aligning linguistic complexity with the difficulty of English texts for L2 learners based on CEFR levels. *Studies in Second Language Acquisition*, 1–28. <https://doi.org/10.1017/S0272263125101125>
- Zhi, R., & Derakhshan, A. (2024). Modelling the interplay between resilience, emotion regulation and psychological well-being among Chinese English language teachers: The mediating role of self-efficacy beliefs. *European Journal of Education*, 59(3), 1–9. <https://doi.org/10.1111/ejed.12643>

Diah Kurniati, Ahdi Riyono, Achmad Solaeman, Arceli R. Millan, Nooreen Noordin

Exploring English proficiency levels of Indonesian university students through EnglishScore based assessment