



6th CELTI

Conference in English Language Teaching

AN INTERNATIONAL VIRTUAL CONFERENCE

Decolonize English at the Age of
Artificial Intelligence: Rethinking
Language, Literature, and Pedagogy

Students' Perceptions of EF SET as a Digital Performance-Based English Proficiency Assessment

Author:

Welly Tan

Affiliation:

Universitas PGRI Semarang,
Semarang, Indonesia

Corresponding email:

wellytan13@gmail.com

DOI:

<https://doi.org/10.24090/6jx.gev77>

Pages:

237-246

Abstract: This study explores students' perceptions of the EF Standard English Test (EF SET) as a digital, performance-based English proficiency assessment within an international curriculum setting. As digital language assessment tools become more common, understanding how students view their effectiveness is crucial for evaluating their educational value. However, little research has specifically examined students' perceptions of EF SET in school environments. This study used a quantitative descriptive approach involving 25 Grade 9 students who completed the EF SET. Data were gathered using a structured questionnaire that measured perceived usefulness, ease of use, fairness, and overall assessment experience. Students' proficiency levels were considered in the context of interpreting their responses. Descriptive statistics were used to analyze the data and identify general trends in perceptions. Results show that students generally view EF SET positively, especially regarding accessibility, clear instructions, and immediate feedback. They reported that the test helped them better understand their English proficiency. Still, concerns about fairness and the test's limited ability to assess productive language skills were raised, as the test mainly focused on reading and listening. These findings suggest EF SET is effective as a digital assessment tool, but its role as a performance-based measure is partial. It should be complemented with other assessment methods for a more complete evaluation of students' language skills. Overall, this study highlights the importance of including students' perspectives in implementing digital language assessments in international education settings.

Keywords: Digital Language Assessment; Performance-Based Assessment; EF SET; Student Perceptions; English proficiency

Copyright © 2026 by Author/s. This is an open-access article distributed under the Creative Commons Attribution-Share Alike 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

The rapid progress of digital technology has significantly changed various aspects of education, including how English language testing is conducted. Digital Language Assessment (DLA) has developed as a response to the growing need for flexible, adaptive, and technology-based evaluation systems in modern education. Instead of just converting traditional paper tests to digital formats, DLA broadens assessment by using authentic, ongoing, and data-driven methods through platforms like Computer-Assisted Language Learning (CALL) and Mobile-Assisted Language Learning (MALL). In this model, language assessment is no longer limited to specific language skills but now includes communicative ability in real-world settings

(Crompton et al., 2024; Tuncer & Özkan, 2021). Additionally, digital tools allow for instant feedback, which is important for improving learners' awareness of their language skills. As a result, DLA has become a vital part of modern English language education that meets 21st-century learning needs.

The integration of technology into language assessment is closely linked to theoretical frameworks such as Technology-Enhanced Language Learning (TELL) and Technological Pedagogical Content Knowledge (TPACK), which highlight the interaction among technology, pedagogy, and content knowledge. In this view, technology acts as a support tool rather than an end goal, aiding more effective teaching and assessment methods (Yuyun, 2018). Systematic reviews have shown that online learning environments help improve specific language skills, such as writing, listening, and vocabulary, especially when they include authentic and collaborative learning activities (Ayar, 2021). However, the research also points out that the success of technology use largely depends on how learning and assessment tasks are designed to promote higher-order thinking skills. Without well-planned tasks, technology might only increase accessibility without significantly improving the quality of language-learning outcomes (Ayar, 2021). Thus, assessing the effectiveness of digital assessment tools remains an important issue, particularly from the perspective of students.

In contemporary language assessment practices, Performance-Based Assessment (PBA) has gained prominence for its ability to evaluate learners' skills through authentic, meaningful tasks. Unlike traditional assessments that depend heavily on multiple-choice formats, PBA emphasizes tasks that require students to actively use language in real-life situations, such as presentations, written compositions, and interactive communication (Tuncer & Özkan, 2021). The integration of PBA into digital environments offers new opportunities to use various digital artifacts, including e-portfolios, video recordings, and project-based tasks, which provide a more complete picture of learners' language abilities. Additionally, this approach supports the development of critical thinking, collaboration, and problem-solving skills, essential competencies in the 21st century (Aprianto & Muhlisin, 2022; Lo & Shi, 2024). However, implementing PBA also involves challenges, especially in ensuring scoring consistency, standardization, and inter-rater reliability, which require clearly defined rubrics and strong assessment frameworks.

As digital assessment tools become more common, understanding students' acceptance and perceptions of these technologies is essential. The Technology Acceptance Model (TAM) offers a useful theoretical perspective for exploring how users view and adopt digital systems. TAM emphasizes two main factors, perceived usefulness and perceived ease of use, that shape users' attitudes and behavioural intentions toward using technology (Allen, 1999). In digital language testing, these factors affect how comfortable, motivated, and confident students feel when taking online assessments. Past research shows that elements like accessibility, clear instructions, and immediate feedback greatly improve students' acceptance of digital testing platforms (Crompton et al., 2024). Still, issues such as technological infrastructure, digital inequality, and concerns about assessment accuracy are critical challenges to address when implementing digital assessments.

One prominent example of a digital language assessment tool is the EF Standard English Test (EF SET), an online, freely accessible test that measures English proficiency based on the Common European Framework of Reference for Languages (CEFR). The Common European Framework of Reference for Languages (CEFR) provides a standardized system for evaluating language proficiency across multiple levels. CEFR descriptors allow educators to classify learners' abilities systematically and assess key components such as grammar, vocabulary, and coherence. Research shows that CEFR-based assessment can effectively capture variations in learners' language performance and support outcome-based education frameworks (Wahyuni et al., 2026). EF SET offers several advantages, including high accessibility, flexible test administration, and instant scoring, making it an attractive option for educational institutions and learners worldwide. In many contexts, it is used as a screening tool to quickly assess students' language proficiency levels. Despite these advantages, EF SET shares common limitations with other online assessments, particularly in its focus on receptive skills such as reading and listening, which may not fully capture learners' overall communicative competence (Tuncer & Özkan, 2021; Crompton et al., 2024). As a result, questions arise regarding its effectiveness as a comprehensive performance-based assessment tool.

Beyond technical factors, students' perceptions of digital language assessments are crucial in determining their effectiveness and sustainability. Empirical research shows that students generally

view computer-based language tests positively when they are seen as user-friendly, fair, and transparent (Hussain et al., 2021). These perceptions not only affect students' testing experiences but also influence their motivation, engagement, and confidence in using the target language. However, concerns about fairness often arise, especially when technical issues or environmental factors impact test performance. Additionally, the limited ability of many digital tests to assess productive skills, such as speaking and writing, remains a key concern in evaluating their overall validity. Thus, exploring students' perceptions is vital to gaining a full understanding of the strengths and weaknesses of digital language assessments.

Despite rising research on digital language assessment, a significant gap remains, especially regarding students' perceptions of EF SET in formal educational settings. Most studies focus on technical issues like validity and reliability or on broader topics such as technology integration in language learning. However, empirical research that specifically explores students' experiences and perceptions as primary users of EF SET is still limited. This gap highlights the need for studies adopting a user-centered approach, particularly in international curriculum contexts where digital assessments are increasingly common. Addressing this gap is vital for a deeper understanding of the role and effectiveness of EF SET in education.

The integration of artificial intelligence in education has significantly transformed language learning and assessment by enabling personalized learning, real-time feedback, and adaptive systems (Sethi & Jain, 2024). Furthermore, traditional assessment methods are increasingly criticized for providing only limited snapshots of student performance, highlighting the need for more dynamic digital assessment approaches (Swiecki et al., 2022).

Based on the considerations mentioned above, this study aims to explore students' perceptions of EF SET as a digital, performance-based English proficiency assessment tool. It seeks to identify both the strengths and limitations of EF SET within an educational context and to examine how students' perceptions influence its effectiveness as an assessment instrument. The novelty of this study lies in its student-centered evaluation of a widely used digital assessment tool, incorporating theoretical perspectives from DLA, PBA, and TAM to provide a comprehensive analytical framework. Accordingly, the research question guiding this study is: **How do students perceive EF SET as a digital performance-based English proficiency assessment?**

Literature Review

Digital Language Assessment in Modern English Language Learning

Digital Language Assessment (DLA) has become a vital part of modern English language education, reflecting the broader digital shift in teaching and learning methods. DLA involves assessment processes delivered, managed, and analyzed through digital platforms, enabling more flexible, scalable, and interactive evaluation systems. Unlike traditional assessment methods, which often rely on static, decontextualized test formats, DLA emphasizes authenticity, continuity, and integration with the learning process (Crompton et al., 2024; Tuncer & Özkan, 2021). The rise of technologies such as Computer-Assisted Language Learning (CALL) and Mobile-Assisted Language Learning (MALL) has greatly contributed to the development of digital assessment practices that reflect real-world language use.

From a teaching perspective, DLA is closely connected to frameworks such as Technology-Enhanced Language Learning (TELL) and Technological Pedagogical Content Knowledge (TPACK), which emphasize the importance of integrating technological tools with effective teaching methods and content knowledge (Yuyun, 2018). These frameworks highlight that technology should not just be a means of delivery but should improve the quality of learning and assessment. Empirical studies show that web-based learning environments can enhance students' language skills, especially in writing, listening, and vocabulary development, when supported by authentic and collaborative activities (Ayar, 2021). However, the success of DLA largely depends on how tasks are designed, as poorly constructed digital assessments may not foster higher-order thinking skills and meaningful language use.

Furthermore, DLA provides immediate feedback and ongoing assessment, which are crucial for promoting learners' self-regulated learning (SRL) and metacognitive awareness. The use of

multimedia tools, such as audio, video, and digital portfolios, offers more comprehensive representations of learners' language skills and encourages more engaging assessment methods. However, expanding DLA also raises concerns about equity, accessibility, and the need for strong technological infrastructure, all of which remain significant challenges in its implementation (Crompton et al., 2024; Moorhouse & Kohnke, 2021).

Performance-Based Assessment in Language Learning

Performance-Based Assessment (PBA) signifies a shift in language assessment, emphasizing learners' ability to use language in real-life and meaningful situations. Unlike traditional tests that focus on specific language knowledge, PBA assesses learners' communicative skills through tasks such as presentations, written essays, and interactive conversations (Tuncer & Özkan, 2021). This method supports modern views of language learning that highlight practical language use, pragmatics, and situational communication.

The integration of PBA into digital environments further boosts its potential by allowing the use of various digital artifacts, including e-portfolios, video recordings, and project-based tasks. These tools offer a more comprehensive and holistic view of learners' language skills, capturing both process and product aspects of learning (Aprianto & Muhlisin, 2022; Lo & Shi, 2024). Additionally, digital PBA supports ongoing assessment and iterative feedback, which are crucial for developing higher-order thinking skills and encouraging learner autonomy.

Performance-Based Assessment (PBA) emphasizes evaluating learners' abilities through authentic, meaningful tasks. One widely used approach in PBA is rubric-based assessment, which provides structured criteria for evaluating performance at different levels. Rubrics are considered effective tools for both formative and summative assessment because they include qualitative descriptions of performance criteria and support consistency in scoring (Kan, 2007).

Despite its benefits, implementing PBA faces several challenges. One major concern is maintaining reliability and consistency in scoring, especially when assessments involve subjective judgments. Developing clear, standardized rubrics is essential to address this and ensure fairness across contexts (Tuncer & Özkan, 2021). Additionally, scaling PBA for large assessments remains difficult, particularly when compared to automated testing systems. These limitations emphasize the need for balanced assessment methods that combine PBA's strengths with the efficiency of digital testing.

Technology Acceptance Model (TAM) in Digital Assessment Contexts

The successful implementation of digital language assessments depends not only on their technical quality but also on users' acceptance of the technology. The Technology Acceptance Model (TAM) provides a widely used framework for understanding how users perceive and adopt technological systems. According to TAM, two main factors, perceived usefulness and perceived ease of use, shape users' attitudes toward technology and their intention to use it (Allen, 1999).

In digital language testing, perceived usefulness refers to how much students believe the assessment improves their understanding of their language skills or helps their learning. Perceived ease of use, however, describes how simple and user-friendly the assessment platform is. Research shows that digital assessments that are accessible, well-organized, and give immediate feedback are more likely to be viewed positively by students (Crompton et al., 2024). These factors lead to greater engagement, less anxiety, and better overall testing experiences.

Digital technologies also support creativity and higher-order thinking skills in educational settings by providing tools for idea development, collaboration, and the creation of digital products (Bereczki & Kárpáti, 2021).

However, the adoption of digital assessments is also influenced by external factors such as technological infrastructure, digital literacy, and institutional support. Challenges like unstable internet connections, unfamiliar interfaces, and concerns about data security can negatively impact students' perceptions and acceptance of digital testing systems (Moorhouse & Kohnke, 2021). Therefore, understanding students' perceptions through the lens of TAM is essential for designing effective and user-centered digital assessments.

Validity, Reliability, and Limitations of Online English Proficiency Tests

Validity and reliability are essential principles in language assessment, determining how well a test measures what it intends to and yields consistent results. In the context of online English proficiency tests, these concepts become more complex due to technological factors. Validity in digital assessments includes not only construct validity but also ecological validity, which refers to how well test tasks mirror real-world language use (Crompton et al., 2024; Sharadgah & Sa'di, 2022).

Reliability in online assessments can be influenced by several factors, such as technical issues, differences in testing environments, and variations in device quality. Studies on Computer-Assisted Language Testing (CALT) have indicated that although students generally view digital tests favorably, concerns about technical reliability and consistency persist. These issues emphasize the need for a stable technological infrastructure and standardized testing conditions.

The findings of this study support previous research suggesting that traditional assessment approaches often provide only limited snapshots of learners' abilities and may not fully capture the complexity of learning processes (Swiecki et al., 2022).

The EF Standard English Test (EF SET), as an example of an online proficiency test, offers benefits including accessibility, affordability, and alignment with the CEFR framework. However, it also shares common limitations with other digital assessments, particularly in its emphasis on receptive skills. This limitation raises questions about its ability to fully measure learners' communicative competence, particularly in productive skills such as speaking and writing (Tuncer & Özkan, 2021; Crompton et al., 2024). As a result, EF SET is often seen more as a screening tool rather than a comprehensive assessment instrument.

Students' Perceptions of Digital English Language Assessment

Students' perceptions are vital in shaping the effectiveness and acceptance of digital language assessments. Favorable perceptions are often linked to factors like ease of use, fairness, accessibility, and immediate feedback. Empirical studies show that students generally view computer-based language testing positively when these factors are present (Hussain et al., 2021). Such perceptions can boost students' motivation, engagement, and confidence, leading to improved learning outcomes.

At the same time, students' perceptions are shaped by various psychological and contextual factors, including test anxiety, familiarity with technology, and perceived relevance of the assessment. Research indicates that digital assessments that include interactive and personalized features, such as adaptive testing and AI-driven feedback, can boost students' engagement and learning experiences (Crompton et al., 2024; Shurovi et al., 2024). However, concerns about fairness, accuracy, and the inability of some digital tests to evaluate productive skills remain important issues.

Furthermore, integrating digital assessment with reflective practices like e-portfolios and self-assessment has been shown to boost students' self-regulated learning and metacognitive awareness (Yuyun, 2018). These findings highlight the importance of designing digital assessments that not only evaluate learning outcomes but also support the learning process itself. Therefore, understanding students' perceptions is crucial for identifying the strengths and limitations of digital assessment tools and for guiding their effective use in educational settings.

Research Method

This study used a quantitative descriptive research design to examine students' perceptions of the EF Standard English Test (EF SET) as a digital, performance-based English proficiency assessment. The quantitative descriptive method was chosen because it allows for the systematic collection and analysis of numerical data to identify patterns, trends, and overall

tendencies in participants' responses. This design is particularly well-suited to studies exploring perceptions, as it enables researchers to quantify subjective experiences and interpret them using statistical analysis. Additionally, this approach aligns with previous studies in digital language assessment that highlight the importance of capturing learners' responses to technology-mediated testing environments (Crompton et al., 2024; Hussain et al., 2021).

The participants in this study included 25 Grade 9 students enrolled in an international curriculum. The selection was based on purposive sampling, as all students had prior experience with digital learning platforms and were required to complete the EF SET as part of their academic activities. This ensured that participants were familiar with the digital testing environment, which is essential for obtaining valid responses related to perceived usefulness and ease of use, as highlighted in the Technology Acceptance Model (TAM) framework (Allen, 1999). Additionally, considering students' proficiency levels provided important insight into how learners with varying levels of English ability perceive digital assessment tools.

The primary tool used for data collection was a structured questionnaire designed to measure four key constructs: perceived usefulness, perceived ease of use, perceived fairness, and overall assessment experience. These constructs were based on the Technology Acceptance Model and expanded to include dimensions specific to digital language testing. The questionnaire used Likert-scale items, enabling respondents to express their level of agreement with each statement. Using a structured questionnaire ensures consistency in data collection and supports statistical analysis. The instrument was designed to capture essential features of effective digital assessment, including accessibility, clarity of instructions, and feedback mechanisms (Crompton et al., 2024).

Data collection took place after students completed the EF SET under controlled classroom conditions to reduce external factors that could influence their responses. Before giving the questionnaire, students received a brief explanation of the study's purpose and were assured of the confidentiality and anonymity of their responses. This step was crucial for reducing response bias and promoting honest feedback. The EF SET was administered following standard online testing procedures, allowing students to access the test on digital devices with internet access. This approach reflects common practices in digital language assessment and ensures ecological validity, as students completed the test in an environment similar to real-world usage contexts (Tuncer & Omlkan, 2021).

The data collected from the questionnaire were analyzed using descriptive statistics, including mean scores, percentages, and frequency distributions. These approaches helped identify overall trends in students' perceptions across the measured constructs. Descriptive statistics are especially appropriate for this study because the main goal is to describe and interpret patterns in students' responses rather than to determine cause-and-effect relationships. The analysis also accounted for students' proficiency levels to gain a more detailed understanding of how perceptions might differ across groups of learners.

To ensure the rigor of the study, attention was paid to the validity and reliability of the instruments. Content validity was established by aligning the questionnaire items with theoretical constructs from Digital Language Assessment, Performance-Based Assessment, and TAM frameworks. Additionally, the clarity and relevance of the items were reviewed to ensure they accurately reflected students' perceptions of the EF SET. Reliability was maintained by administering the instrument consistently and using standardized response formats. These steps are essential in digital assessment research, where technological and contextual factors can influence data consistency (Hussain et al., 2021).

The use of CEFR levels to interpret students' proficiency is consistent with previous research demonstrating that CEFR descriptors provide a reliable framework for classifying learners' language abilities across proficiency levels (Wahyuni et al., 2026).

Overall, this methodological approach offers a systematic and reproducible framework for analyzing students' perceptions of EF SET. By combining quantitative descriptive analysis with established theoretical frameworks, this study provides a solid foundation for understanding learners' views of digital performance-based assessments in educational settings. The results from this method are expected to contribute to the ongoing discussion on the effectiveness and limitations of digital language assessment tools.

Finding (s)

This section presents the study's findings based on the EF Standard English Test (EF SET) results, along with students' responses to the questionnaire and interview data. The findings are organized into three main parts: (1) overall EF SET performance, (2) analysis of language skill components, and (3) students' perceptions of the EF SET, including perceived usefulness, ease of use, and perceived validity. The data are presented descriptively without interpretation.

Overall EF SET Performance

The initial analysis examined students' overall scores on the EF SET. The results show that students' scores ranged from 44 to 85, with an average score of 64.24. These scores were then classified based on the Common European Framework of Reference for Languages (CEFR) levels.

Table 1. Distribution of English Proficiency Levels Based on CEFR

CEFR Level	Category	Number of Students	Percentage
C2	Proficient	6	24%
C1	Advanced	10	40%
B2	Upper-Intermediate	8	32%
B1	Intermediate	1	4%
Total		25	100%

The data indicate that most students showed relatively high English skills. A total of 16 students (64%) were classified as advanced to proficient levels (C1-C2). Additionally, 8 students (32%) were at the upper-intermediate level (B2), while only 1 student (4%) was at the intermediate level (B1). The distribution shows that most participants had a solid foundation in English skills, mainly at higher CEFR levels. However, the presence of students at different proficiency levels also highlights variability within the group.

Analysis of Language Skill Components

In addition to overall scores, the EF SET results provided insights into students' performance across different language skill components. The data reveal differences in students' performance across these areas. For example, a student identified by the initials V.I. scored an overall 85, which corresponds to the C2 (Proficient) level. According to the test report, this student showed consistently high performance across assessed components, especially in reading and writing, where all indicators were rated at the highest level. In contrast, some students exhibited noticeable gaps between different language skill areas. One example is the student with the initials T.S., who scored an overall 73, classified within the C2 level. However, this student achieved a score of 38 in the speaking component, which corresponds to the A2 level. This variation shows that students' performance across different language skills isn't always uniform. Some students performed well in receptive skills, such as reading and listening, but had lower scores in productive skills, particularly speaking. The data also show that such discrepancies weren't isolated but observed among several participants. This suggests that while overall proficiency levels may look high, individual skill areas can vary widely among learners.

Discussion

This study aimed to examine students' perceptions of the EF Standard English Test (EF SET) as a digital, performance-based English proficiency assessment. The findings highlight several important patterns regarding students' proficiency levels, their experiences with the digital assessment platform, and their views on its usefulness, usability, and validity. These results are discussed in relation to existing literature on Digital Language Assessment (DLA), Performance-Based Assessment (PBA), and the Technology Acceptance Model (TAM), offering a comprehensive understanding of the EF SET's role

First, the results show that most students achieved relatively high proficiency levels, with 64% of participants classified as advanced to proficient (C1–C2). This indicates that the participants had a strong foundation in English, which could influence how they perceive and interact with digital assessment tools. According to the literature, learners with higher proficiency levels tend to engage more effectively with digital platforms due to their stronger linguistic skills and greater familiarity with academic tasks (Crompton et al., 2024). In the context of Digital Language Assessment, this proficiency distribution is important because it highlights the potential of digital tools to support learners who are already at higher levels of language ability. At the same time, the variation in participants' scores suggests that digital assessments should accommodate different proficiency levels, aligning with the principles of adaptive and scalable assessment systems within DLA frameworks (Tuncer & Özkan, 2021).

The analysis of language skill components further shows that students' performance varies across different skills. The observed differences between receptive skills (such as reading and listening) and productive skills (such as speaking) highlight an important limitation of EF SET as a performance-based assessment tool. For example, the case of student T.S., who received a high overall score but a much lower speaking score, illustrates the uneven distribution of language abilities within individual learners. This finding aligns with the literature on Performance-Based Assessment, which emphasizes that language proficiency is multidimensional and cannot be fully captured through a limited set of skill areas (Tuncer & Özkan, 2021). PBA supports the inclusion of authentic tasks that assess both receptive and productive skills, providing a comprehensive evaluation of learners' communicative competence.

In digital assessment settings, capturing productive skills remains a significant challenge. While EF SET offers quick and standardized measurement of reading and listening, it does not fully align with the principles of PBA, which require learners to demonstrate language use in authentic, context-rich tasks (Aprianto & Muhlisin, 2022; Lo & Shi, 2024). The results of this study confirm this limitation, as students themselves recognized the gap between the skills assessed by EF SET and their overall communicative abilities. This indicates that EF SET works better as a screening or placement tool rather than as a comprehensive performance-based assessment.

From the perspective of the Technology Acceptance Model (TAM), the findings indicate that students generally have positive perceptions of EF SET in terms of perceived usefulness and perceived ease of use. Most students reported that the test helped them understand their current level of English proficiency and identify areas for improvement. This supports the idea of perceived usefulness, which refers to the extent to which users believe a technology improves their performance or results (Allen, 1999). The immediate feedback provided by EF SET seems to play a key role in shaping this perception, as it allows students to quickly access information about their proficiency levels.

Furthermore, students reported that the platform was easy to use, with clear instructions and few technical issues. This demonstrates the concept of perceived ease of use, which is a crucial factor in technology acceptance. According to the literature, user-friendly, accessible digital tools are more likely to be adopted and viewed positively by learners (Crompton et al., 2024). The positive feedback from students in this study indicates that EF SET effectively meets these standards, leading to a positive user experience.

The importance of usability in shaping students' perceptions is especially significant in digital assessment settings. As previous studies have shown, technical barriers such as complicated interfaces or unreliable systems can negatively affect students' experiences and reduce their engagement with digital assessments (Moorhouse & Kohnke, 2021). Conversely, this study's findings suggest that EF SET provides a smooth testing experience, boosting students' comfort and confidence with the platform. This emphasizes that usability is a vital factor for the successful implementation of digital language assessments.

Beyond usefulness and ease of use, students' perceptions of fairness and validity also emerged as key themes in this study. While most students found EF SET to be a helpful and accessible tool, some participants expressed concerns about its ability to fully capture their language skills. Specifically, students pointed out the lack of speaking and writing components as a limitation of the test. This

concern is closely tied to the concept of validity in language assessment, which refers to how well a test measures the intended construct (Crompton et al., 2024; Sharadgah & Sa'di, 2022).

The findings indicate that EF SET shows partial validity: it effectively assesses certain aspects of language proficiency, especially receptive skills, but does not cover the full scope of communicative competence. This agrees with previous research suggesting that online language tests often emphasize efficiency and scalability over comprehensiveness (Hussain et al., 2021). Therefore, while these tests can offer useful insights into learners' abilities, they might not be enough as standalone assessment tools in educational settings.

The issue of ecological validity is also relevant here. Performance-Based Assessment highlights the importance of evaluating language use in real-world settings, which often involve interactive and productive tasks. However, EF SET's focus on reading and listening limits its ability to reflect genuine language use. The students' responses in this study emphasize this limitation, as they recognized that their speaking and writing skills were not fully represented in the test results. This finding underscores the need for more comprehensive assessment methods that incorporate multiple skill areas.

Another important aspect of the findings is the connection between students' perceptions and their learning experiences. The positive perceptions of EF SET, especially regarding feedback and usability, suggest that digital assessments can support language learning. Immediate feedback, in particular, is a key factor in encouraging self-regulated learning (SRL), as it helps learners reflect on their performance and identify areas for improvement (Crompton et al., 2024; Shurovi et al., 2024). The students' responses in this study show that EF SET contributes to this by providing clear, accessible information about their proficiency levels.

However, the limitations identified by students also highlight the need for additional assessment methods. As noted in the literature, combining digital assessments with performance-based tasks such as presentations, portfolios, and interactive activities can offer a more comprehensive evaluation of learners' abilities (Tuncer & Özkan, 2021; Aprianto & Muhlisin, 2022). This integrated approach aligns with the principles of DLA and PBA, which stress the importance of using multiple assessment methods to reflect the complexity of language learning.

This study's findings add to the field of language assessment by providing empirical evidence on students' perceptions of EF SET in an educational setting. By emphasizing the learner's perspective, it addresses a gap in existing research, as most prior studies focus on the technical aspects of digital assessments rather than on user experiences. The use of DLA, PBA, and TAM frameworks offers a thorough analytical approach for understanding how learners perceive and use digital assessment tools.

Regarding implications, the findings indicate that educators and institutions should weigh both the strengths and limitations of EF SET when incorporating it into their assessment methods. Although EF SET is a useful digital tool for gauging certain language skills, it should not be the only way to measure students' abilities. Instead, it should be combined with performance-based assessments that evaluate productive skills to give a more complete picture of learners' competence.

Additionally, the findings emphasize the need for designing digital assessments that are not only technically efficient but also pedagogically meaningful. This involves ensuring assessments align with learning objectives, offer meaningful feedback, and support students' learning processes. Incorporating advanced technologies, such as artificial intelligence and adaptive testing, can further improve the effectiveness of digital assessments, but these innovations must be carefully evaluated to ensure validity, fairness, and reliability (Crompton et al., 2024; Sharadgah & Sa'di, 2022).

Although this study makes valuable contributions, it also has some limitations. The sample size was relatively small and confined to a specific educational setting, which could limit the generalizability of the results. Furthermore, the study primarily relied on self-reported data, which may be subject to response bias. Future research can address these limitations by using larger, more diverse samples and incorporating additional data sources, such as performance assessments and longitudinal data.

In conclusion, this study's findings highlight both the strengths and limitations of EF SET as a digital, performance-based English proficiency test. Although the test is generally seen as useful,

accessible, and easy to use, its limited ability to evaluate productive skills reduces its effectiveness as a comprehensive assessment tool. These results emphasize the need for a balanced approach to language assessment that combines digital tools with performance-based methods, providing a more complete and accurate evaluation of learners' language skills.

Conclusion

This study examined students' perceptions of the EF Standard English Test (EF SET) as a digital, performance-based English proficiency assessment within an educational setting. The results show that EF SET is generally viewed favorably by students, particularly for its accessibility, ease of use, and ability to provide immediate feedback. Students reported that the test helps them understand their current English proficiency level and identify areas for improvement, highlighting its usefulness as a diagnostic and self-assessment tool.

Additionally, the results showed that most participants attained relatively high proficiency levels, with many students falling into the advanced to proficient CEFR categories. This indicates that EF SET can effectively distinguish learners' proficiency levels in a digital testing environment. However, the analysis of language skill components revealed differences in students' performance across skills, emphasizing that language proficiency is inconsistent and may vary between receptive and productive abilities.

Despite its strengths, the study also identified important limitations of EF SET as a performance-based assessment tool. Students expressed concerns about its limited scope, especially its focus on reading and listening skills, which do not fully reflect their overall communicative competence. This shows that while EF SET is effective as a digital assessment tool, its role as a performance-based assessment is only partial and needs to be complemented by other methods that evaluate productive skills such as speaking and writing.

This study's findings contribute to digital language assessment by highlighting the importance of including students' perspectives when evaluating assessment tools. By combining insights from the Digital Language Assessment, Performance-Based Assessment, and Technology Acceptance Model frameworks, this approach offers a comprehensive view of how learners experience digital assessments.

For future research, it is advisable to include larger, more diverse samples to improve the generalizability of the results. Additionally, future studies could look into integrating EF SET with performance-based assessment methods to provide a more comprehensive evaluation of language proficiency. Further exploration of emerging technologies, such as adaptive testing and AI-driven assessments, may also yield valuable insights into enhancing the effectiveness and validity of digital language evaluations.

References

- Adinolou, N. A., & Far, L. M. (2014). The relationship of self-efficacy beliefs, writing strategies, and the correct use of conjunctions in Iranian EFL learners. *International Journal of Applied Linguistics & English Literature*, 3(4), 221-227. <http://dx.doi.org/10.7575/aiac.ijalel.v.3n.4p.221>
- Arabski, J., & Wojtaszek, A. (Eds.), (2011). *Aspects of culture in second language acquisition and foreign language learning*. Berlin: Springer.
- Atmazaki, Ali, N. B. V., Muldian, W., Miftahussururi, Hanifah, N., Nento, M. N., & Akbari, Q. S. (2017). *Panduan gerakan literasi nasional [National literacy movement guidelines]*. Jakarta: Kementerian Pendidikan dan Kebudayaan Republik Indonesia
- Aunurrahman, Hamied, F., & Emilia, E. (2017). Realizing a good education in an Indonesian university context. In A. G. Abdullah, I. Hamidah, S. Aisyah, A. A. Danuwijaya, G. Yuliani, & H. S. H. Munawaroh (Eds.), *Ideas for 21st Century Education: Proceedings of the Asian Education Symposium (AES 2016)* (pp. 297-300). London: Routledge. <https://doi.org/10.1201/9781315166575>
- Allen, B. F. H. (1999). *Teaching English as a Second or Foreign Language: An Annotated Bibliography of*

- Journals in the Field. *Collection Building*, 18(1), 16–31. <https://doi.org/10.1108/01604959910256517>
- Aprianto, D., & Muhlisin, M. (2022). The Use of Self-Recording Videos as ESL Learning Method in Increasing English Learning Performance. *Empiricism Journal*, 3(2), 246–253. <https://doi.org/10.36312/ej.v3i2.993>
- Ayar, Z. (2021). The Systematic Review of Studies About Web-Based Learning in Language Teaching Context. *E-KaQkas Eğitim Araştırmaları Dergisi*, 8(2), 123–136. <https://doi.org/10.30900/ka@kasegt.957362>
- Bereczki, E. O., & Kárpáti, A. (2021). Technology-enhanced creativity: A multiple case study of digital technology-integration expert teachers' beliefs and practices. *Thinking Skills and Creativity*, 39(January). <https://doi.org/10.1016/j.tsc.2021.100791>
- 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/10.1111/bjet.13460>
- Hussain, R., Zahid, M., & Alam, Y. (2021). Ensuring Viable Solutions in Resource-Constrained Language Teaching Settings: ESL Learners Perceptions About Computer-Assisted Language Testing (CALT). *Sustainable Business and Society in Emerging Economies*, 3(4), 509–519. <https://doi.org/10.26710/sbsee.v3i4.2052>
- Kan, A. (2007). An Alternative Method in the New Educational Program from the Point of Performance-based Assessment: Rubric Scoring Scales. *Educational Sciences: Theory & Practice*, 7(January 2007), 144–152.
- Lo, N., & Shi, H. (2024). The Perceptions of Undergraduate Students Toward Reading Contemporary Fiction in English: A Case Study of Content-Based ESL Instruction at a Self-Financed Tertiary Institution in Hong Kong. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1395168>
- Moorhouse, B. L., & Kohnke, L. (2021). Responses of the English-Language-Teaching Community to the COVID-19 Pandemic. *Relc Journal*, 52(3), 359–378. <https://doi.org/10.1177/00336882211053052>
- Sethi, S. S., & Jain, K. (2024). AI technologies for social-emotional learning: recent research and future directions. *Journal of Research in Innovative Teaching and Learning*, 17(2), 213–225. <https://doi.org/10.1108/JRIT-03-2024-0073>
- Sharadgah, T. A., & Sa'di, R. A. (2022). A Systematic Review of Research on the Use of Artificial Intelligence in English Language Teaching and Learning (2015-2021): What Are the Current Effects? *Journal of Information Technology Education Research*, 21, 337–377. <https://doi.org/10.28945/4999>
- Shurovi, M., Yahaya, M. F., Hajimia, H., & Hasan, Md. K. (2024). Five Decades of Achievement Motivation Research in ELT: A Systematic Literature Review. *Malaysian Journal of Learning and Instruction*, 21(1), 285–318. <https://doi.org/10.32890/mjli2024.21.1.10>
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3(April). <https://doi.org/10.1016/j.caeai.2022.100075>
- Tuncer, H., & Özkan, Y. (2021). Reflective Tools in an English Language Teaching Context: Contributions, Challenges and Recommendations. *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, 18(1), 181–204. <https://doi.org/10.33437/ksusbd.671482>
- Wahyuni, D., Aini, S. N., & Senowarsito, S. (2026). *Profiling STEM Students' English Writing Skills Based on CEFR Descriptors within an Outcome-Based Education Framework*. (Icesre 2025).
- Yuyun, I. (2018). Curriculum and Technology Design: A Course to Explore Technology Applications in EFL Curriculum Design. *Journal of ELT Research*, 3(1), 78. https://doi.org/10.22236/jer_vol3issue1pp78-86