International Journal of Pedagogy and Education Research

English Receptive-Skills Assessment in Vocational High Schools: Developing a Test Item Bank

Edward[1]; Sorta Hutahaean[2]; Nadila Mudea Najamuddin[3]*
[1,2,3] Department of English Literature, Universitas Lancang Kuning, Pekanbaru, Indonesia
edwards07@unilak.ac.id; sortahutahaean@unilak.ac.id; nadilam@unilak.ac.id*

(*) Corresponding Author

Received: 8 October 2025; Revised: 05 November 2025 ; Accepted: 07 November 2025

ABSTRACT

Assessment of English receptive skills plays a crucial role in vocational high schools, as reading and listening competence underpins students' ability to comprehend functional texts relevant to academic and workplace contexts. However, preliminary analysis revealed that existing assessment instruments did not adequately represent syllabus content, lacked listening components, and had not undergone empirical validation. This study aimed to develop a valid and reliable English receptive-skills test item bank aligned with curriculum demands. Employing a Research and Development design, the study integrated the Define–Design–Develop model with systematic procedures of syllabus analysis, item construction, expert validation, revision, and field trials. Data were collected through document analysis, expert judgment, and students' test results, and analyzed using content validity evaluation, KR-21 reliability estimation, and item difficulty indices. The findings indicate that the developed item bank comprehensively covers syllabus indicators, integrates reading and listening skills, and demonstrates acceptable to high reliability, with most items falling within the medium difficulty range. These results confirm that systematic development and empirical testing significantly enhance assessment quality. In conclusion, the test item bank is suitable for use in semester examinations to ensure fair, accurate, and curriculum-based evaluation. It is recommended that schools institutionalize item banking practices and conduct regular item analysis, while future studies may expand assessment development to productive language skills.

Key Words: Assessment Validity; English Receptive Skills; Item Bank Development; Vocational High School

ABSTRAK

Penilaian keterampilan reseptif bahasa Inggris memainkan peran penting di sekolah menengah kejuruan, karena kompetensi membaca dan menyimak mendukung kemampuan siswa untuk memahami teks fungsional yang relevan dengan konteks akademis dan tempat kerja. Namun, analisis awal menunjukkan bahwa instrumen penilaian yang ada tidak cukup mewakili isi silabus, tidak memiliki komponen menyimak, dan belum melalui validasi empiris. Penelitian ini bertujuan untuk mengembangkan bank soal tes kemampuan bahasa Inggris yang valid dan reliabel yang sesuai dengan tuntutan kurikulum. Dengan menggunakan desain Penelitian dan Pengembangan, penelitian ini mengintegrasikan model Define-Design-Develop dengan prosedur sistematis analisis silabus,

* Corresponding author

konstruksi butir soal, validasi ahli, revisi, dan uji coba lapangan. Data dikumpulkan melalui analisis dokumen, penilaian ahli, dan hasil tes siswa, dan dianalisis dengan menggunakan evaluasi validitas isi, estimasi reliabilitas KR-21, dan indeks kesukaran butir soal. Temuan menunjukkan bahwa bank soal yang dikembangkan secara komprehensif mencakup indikator silabus, mengintegrasikan keterampilan membaca dan menyimak, dan menunjukkan reliabilitas yang dapat diterima hingga tinggi, dengan sebagian besar butir soal termasuk dalam kisaran tingkat kesulitan sedang. Hasil ini menegaskan bahwa pengembangan sistematis dan pengujian empiris secara signifikan meningkatkan kualitas penilaian. Kesimpulannya, bank soal tes ini cocok untuk digunakan dalam ujian semester untuk memastikan evaluasi yang adil, akurat, dan berbasis kurikulum. Disarankan agar sekolah melembagakan praktik bank soal dan melakukan analisis butir soal secara teratur, sementara penelitian di masa depan dapat memperluas pengembangan penilaian untuk keterampilan bahasa produktif.

Kata Kunci: Validitas Penilaian; Kemampuan Reseptif Bahasa Inggris; Pengembangan Bank Soal; Sekolah Menengah Kejuruan

How to Cite: Edward, E; Hutahaean, S & Najamuddin, N. M. (2025). English Receptive-Skills Assessment in Vocational High Schools: Developing a Test Item Bank. IJPER (Indonesian Journal of Pedagogy and Education Research), Vol (2) No (2), pages 60-72. doi: ...

INTRODUCTION

English assessment in vocational high schools has a strategic role in ensuring that students achieve the learning outcomes mandated by the Merdeka Curriculum, particularly in relation to receptive skills, namely reading and listening. As an outcome-based curriculum, the Merdeka Curriculum emphasizes authentic assessment that reflects students' actual language competence rather than rote knowledge. In this context, assessment functions not merely as a tool for grading but as an empirical mechanism for measuring how well students comprehend spoken and written English texts that are relevant to vocational and workplace contexts. Without valid and reliable assessment instruments, teachers may fail to capture students' real proficiency levels, which can lead to misaligned instructional decisions and ineffective learning interventions (Hartell & Buckley, 2021). Therefore, English assessment in vocational education must be systematically designed to reflect curriculum demands, learner needs, and real-world language use.

Receptive skills are widely recognized in applied linguistics as the foundation of language acquisition and communicative competence. Reading and listening provide the primary linguistic input that enables learners to internalize vocabulary, grammatical patterns, and discourse structures (Zhang & Zhang, 2021). In vocational high schools, where students are expected to engage with manuals, instructions, announcements, and workplace-related texts, strong receptive skills are essential for academic success and future employability. Chung & Wan (2025) emphasize that learners who develop effective receptive strategies are better equipped to process meaning, infer context, and respond appropriately in communicative situations. Consequently, assessing receptive skills is not optional but fundamental to understanding how learners access and interpret language input in both educational and professional domains.

Moreover, receptive skills play a critical role in supporting students' comprehension of functional texts, which constitute a core component of English instruction in vocational education. Functional texts such as announcements, instructions, invitations, reports, and short dialogues require learners to understand social functions, text structures, and linguistic features in context (Liu & Chen, 2022). Through reading and listening, students learn how language operates in authentic situations, enabling them to bridge classroom learning with

* Corresponding author

real-life communication needs. Empirical studies have shown that students' ability to comprehend functional texts is closely linked to their receptive proficiency and the quality of assessment practices used to measure it (Hollister et al., 2022). When assessment tasks accurately reflect these text types, they provide meaningful evidence of students' readiness to use English in vocational and workplace settings.

In line with current assessment theory, effective evaluation of receptive skills must be grounded in established principles of validity, reliability, and authenticity. Taye & Mengesha (2024) argue that language tests should represent real language use and align with instructional objectives to ensure meaningful interpretation of results. Within the Merdeka Curriculum framework, this principle becomes even more relevant, as assessment is expected to support differentiated learning and continuous improvement. By focusing on reading and listening comprehension through well-constructed assessment instruments, vocational high schools can ensure that English learning outcomes are empirically measured, pedagogically relevant, and responsive to contemporary educational demands. Thus, the assessment of receptive skills stands as a cornerstone in developing competent, adaptive, and workplace-ready graduates.

Preliminary analysis conducted at SMK N 5 Pekanbaru indicates that the existing English test item bank does not yet function as a comprehensive and valid instrument for measuring students' receptive language skills. One of the most salient weaknesses identified is the incomplete coverage of syllabus-mandated materials, particularly topics such as self-introduction and the use of simple past and present perfect tenses. According to the principles of curriculum-based assessment, evaluation instruments must representatively sample all instructional objectives to ensure fairness and accuracy (Romadin et al., 2022). When certain competencies outlined in the syllabus are excluded from assessment, students' achievement cannot be measured holistically, and the resulting test scores risk misrepresenting actual language proficiency. This gap suggests a misalignment between instructional goals and evaluative practices, which may undermine the effectiveness of semester examinations as tools for academic decision-making.

In addition, the analysis reveals a disproportionate reliance on reading-based tasks, with listening skills entirely absent from the test item bank. This condition contradicts established theories of language assessment that emphasize the integral role of receptive skills such as listening and reading which act as foundational components of language competence (Byram et al., 2013). Listening, as a primary source of linguistic input, is essential for language acquisition and communicative competence, particularly in English as a Foreign Language (EFL) contexts. The exclusion of listening assessment not only limits the scope of evaluation but also reduces the authenticity of the test, as real-life language use typically requires the integration of multiple skills. Consequently, the overemphasis on reading tasks may lead to an imbalanced assessment framework that fails to capture students' actual communicative abilities.

Furthermore, the existing test items demonstrate a weak alignment with key assessment indicators, including social function, text characteristics, and grammatical accuracy. These indicators are explicitly stated in the national curriculum and are fundamental to assessing students' understanding of functional texts in meaningful contexts. Scholars argue that well-designed language tests should measure not only surface-level comprehension but also learners' ability to interpret purpose, structure, and linguistic features of texts (Shoda & Yamanaka, 2022). The lack of systematic integration of these indicators suggests that the test items tend to focus on isolated comprehension questions rather than on a coherent representation of communicative language use. As a result, the assessment may fail

* Corresponding author

to reflect higher-order comprehension and pragmatic understanding, which are critical outcomes of EFL instruction.

Another critical weakness identified in the preliminary analysis is the absence of systematic item analysis, particularly in terms of validity and reliability. Empirical assessment research underscores that test items must be subjected to rigorous analysis to ensure they consistently and accurately measure intended constructs (Arikunto, 2021). Without procedures such as content validation, reliability estimation, and item difficulty analysis, the quality of test items remains unknown and potentially flawed. This lack of empirical evaluation increases the risk of biased, ambiguous, or ineffective items being reused in high-stakes assessments. In the long term, such practices may compromise the credibility of school-based evaluation systems and hinder efforts to improve instructional quality through data-driven decision-making.

Taken together, these findings highlight an urgent need for the systematic development and validation of an English test item bank at SMK N 5 Pekanbaru. Addressing issues of content coverage, skill balance, indicator alignment, and empirical item analysis is essential to ensure that assessment practices are theoretically grounded, curriculum-aligned, and pedagogically sound. By adopting established frameworks in language testing and educational measurement, schools can enhance the quality of their evaluation instruments and provide more accurate evidence of students' English proficiency in vocational education contexts.

The quality of semester examinations is closely tied to the quality of the assessment instruments used to measure students' learning outcomes. When test item banks are poorly constructed, misaligned with the syllabus, or lack systematic validation, the accuracy of the resulting scores becomes questionable. In the context of English language assessment, particularly for receptive skills, such weaknesses may lead to inaccurate representations of students' actual abilities. Hutchins et al., (2024) emphasize that assessment results are only meaningful when test items genuinely measure the intended constructs. If certain competencies outlined in the curriculum are omitted or overrepresented, students may be unfairly advantaged or disadvantaged, thereby compromising the principle of equity in educational assessment. Consequently, semester examinations that rely on inadequately developed test items risk producing biased outcomes that do not reflect learners' true proficiency levels.

Moreover, fairness in assessment is a fundamental principle in educational evaluation, as it ensures that all students are assessed under comparable conditions using instruments of equivalent quality. Widiyanti (2024) argues that fairness is closely related to test usefulness, which includes validity, reliability, and impact. When examination items have never undergone item analysis, such as difficulty and reliability testing, the consistency of scores across different groups and administrations cannot be guaranteed. This situation may result in inconsistent grading practices and undermine stakeholders' trust in the assessment system. In vocational high schools, where assessment outcomes often influence academic progression and future career pathways, the consequences of unfair or inaccurate testing can be particularly significant.

In response to these challenges, this study was designed to develop a test item bank that meets empirical and theoretical standards of educational measurement. Specifically, the study aimed to produce a valid and reliable test item bank for assessing receptive skills at the eleventh grade of SMK N 5 Pekanbaru. The development process was grounded in the principles of curriculum-based assessment, ensuring that each test item aligns with the competency standards and basic competencies specified in the national syllabus. According

* Corresponding author
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

to (Roebianto et al., 2023), content validity is achieved when test items representatively sample the skills and knowledge domains outlined in instructional objectives. By systematically mapping test items to syllabus indicators, this study sought to strengthen the content validity of the semester examination.

Furthermore, the development of a structured test item bank is supported by established theories of assessment and test construction. Herwin & Nurhayati (2021) explain that an item bank is not merely a collection of questions, but a systematically organized set of validated items with known characteristics, such as difficulty level and reliability. Through expert validation and empirical try-outs, the present study ensured that the developed test items functioned as intended when administered to students. This approach aligns with Sudijono's (2011) view that empirical testing is essential to confirm the reliability and practicality of assessment instruments.

Ultimately, the development of a syllabus-aligned test item bank for receptive skills is expected to enhance the accuracy, fairness, and credibility of semester examinations at SMK N 5 Pekanbaru. By providing teachers with a pool of empirically tested and theoretically grounded test items, the assessment process becomes more systematic and defensible. In line with contemporary assessment practices, this study contributes to the improvement of school-based evaluation by promoting evidence-based test development that reflects both curricular demands and students' actual language competencies.

METHODS

Research design

This study adopted a Research and Development (R&D) design to systematically produce and validate an English test item bank focusing on students' receptive skills. R&D was selected because it enables researchers not only to examine existing educational problems but also to generate a practical product that addresses those problems through iterative testing and refinement. According to (Gall et al., 2007), R&D is particularly appropriate for developing educational instruments that require empirical validation before implementation in instructional settings.

To ensure methodological rigor, the study integrated the Thiagarajan Define–Design–Develop (3D) model with Sugiyono's (2017) educational development procedures. The Define stage emphasized a thorough analysis of curriculum demands and existing assessment practices, while the Design stage focused on constructing test specifications and item formats aligned with learning indicators. The Develop stage involved expert validation, revision, and field trials to establish psychometric quality. This hybrid model was chosen because it allows flexibility while maintaining a clear developmental structure, which is essential in test construction research.

The development process was conducted through sequential stages: (1) analyzing the existing test item bank and syllabus, (2) identifying content gaps and assessment weaknesses, (3) designing and developing multiple-choice test items for reading and listening skills, (4) validating the items through expert judgment, and (5) conducting two cycles of revision and piloting. This systematic progression ensured that the final product was theoretically sound, empirically tested, and aligned with curriculum objectives.

Research site and participants

The research was conducted at SMK N 5 Pekanbaru, a vocational senior high school implementing the 2013 Curriculum, where English assessment emphasizes functional texts

* Corresponding author
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

and communicative competence. This site was selected based on preliminary findings indicating misalignment between the syllabus and existing English test items, particularly in assessing listening skills and grammatical indicators. Such contextual relevance strengthened the empirical foundation of the study.

The participants consisted of two distinct groups. First, two English education experts served as validators. They possessed academic and professional experience in language assessment and were responsible for evaluating the test items in terms of content relevance, construction quality, and linguistic accuracy. Expert involvement is critical in test development studies to ensure content validity and alignment with instructional goals (Siregar et al., 2023).

Second, twenty eleventh-grade students participated in the field trials, with ten students involved in each trial phase. The students were selected from the same grade level but participated in different trial cycles to minimize test familiarity effects. Their responses provided empirical data necessary for analyzing reliability and item difficulty. The use of actual test-takers aligns with best practices in educational measurement, where test quality must be evaluated under real classroom conditions.

Data collection and analysis

Data collection in this study employed multiple sources to enhance validity through triangulation. Documentation was used to examine the existing test item bank and the official English syllabus, enabling the researcher to identify inconsistencies between instructional objectives and assessment content. Expert judgment data were collected through validation checklists focusing on material relevance, item construction, and language use. In addition, students' test results from both trial phases served as quantitative data for psychometric analysis.

Data analysis focused on three key aspects of test quality. First, content validity was examined by evaluating the alignment between test items, syllabus indicators, and learning objectives. This approach follows the principle that a valid test must adequately represent the domain it intends to measure (Widyastuti & Utami, 2018). Second, test reliability was calculated using the KR-21 formula, which is appropriate for multiple-choice tests with dichotomous scoring. Reliability analysis aimed to determine the consistency of test scores across administrations.

Third, item difficulty indices were computed to identify whether test items fell within acceptable difficulty ranges. Items that were too easy or too difficult were revised or discarded, as balanced difficulty is essential for discriminating students' ability levels effectively (Arikunto, 2021). Through these analyses, the study ensured that the developed test item bank met both theoretical and empirical standards of educational measurement.

FINDINGS AND DISCUSSION

Findings

The initial stage of this study involved a systematic analysis of the existing English test item bank used at SMK N 5 Pekanbaru. The findings clearly indicated that the test items did not fully represent the competencies outlined in the syllabus, particularly for receptive skills. Several essential materials mandated by the curriculum, such as self-introduction texts and the use of simple past and present perfect tenses, were absent. Moreover, the assessment focused almost exclusively on reading, neglecting listening skills altogether. This imbalance contradicts the theoretical view that receptive skills should be assessed in an integrated

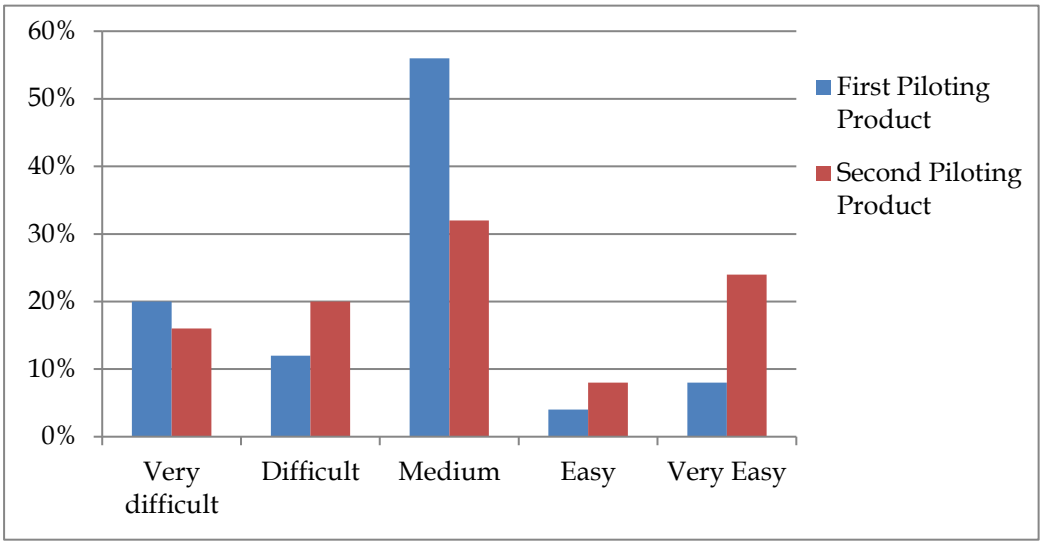
*** Corresponding author**
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

manner because both function as primary channels of language input (Leung & Cheng, 2019). As a result, the original test item bank was insufficient for measuring students' comprehensive receptive language competence.

In response to these shortcomings, a revised test item bank was systematically developed based on the syllabus requirements and established principles of language assessment. The new test item bank covered all required instructional materials, including functional texts for self-introduction and grammatical structures such as simple past and present perfect tense. Importantly, the assessment integrated both reading and listening sections, thereby aligning the test format with communicative language teaching principles that emphasize balanced skill development (Byram et al., 2013). Each test item was carefully designed to assess not only surface-level comprehension but also deeper aspects of language use, including social function, text characteristics, and grammatical accuracy. This comprehensive coverage ensured that the test items reflected real communicative contexts, which is essential for valid language assessment in vocational education settings.

To ensure the quality of the developed instrument, expert validation was conducted involving specialists in English language education and assessment. The validation process focused on content relevance, item construction, and linguistic clarity. After several revisions based on expert feedback, the test items were judged to have met established standards of content and construct validity. This finding is consistent with the view that expert judgment is a critical step in ensuring content validity, particularly when test items are intended to represent specific curriculum objectives (Bhoki, 2025). The validation results confirmed that the revised test item bank was appropriately aligned with the syllabus and suitable for measuring the intended learning outcomes.

Figure 1 Item Difficulty Percentage of Test Item Bank



Empirical analysis on the above figure further demonstrated the improvement in the quality of the test item bank, particularly in terms of reliability. The reliability coefficient

* Corresponding author
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

obtained from the first trial was 0.54, which falls within the medium category, indicating a moderate level of score consistency. After revising the items based on empirical data and expert suggestions, the second trial produced a reliability index of 0.72, categorized as high reliability. This increase indicates that the revised test items yielded more consistent measurement results across test administrations. According to Bachman (2004), such improvement reflects enhanced internal consistency and suggests that the test items functioned more effectively as a unified measurement instrument.

In addition to reliability, item difficulty analysis revealed that the majority of test items fell within the medium difficulty range. This balanced distribution is considered a key characteristic of a good test, as items that are neither too easy nor too difficult are more effective in differentiating students' ability levels (Arikunto, 2009). The presence of predominantly medium-difficulty items suggests that the test item bank was capable of accurately capturing variations in students' receptive skills without causing frustration or disengagement. This empirical evidence supports the claim that the developed test items met psychometric standards commonly recommended in educational measurement literature.

Overall, the findings demonstrate that the developed test item bank was both statistically sound and pedagogically appropriate for classroom use. By aligning test content with the syllabus, integrating reading and listening skills, and ensuring acceptable levels of validity, reliability, and item difficulty, the instrument fulfills the essential criteria of a high-quality language assessment tool. Yet, this research confirms that systematic development and empirical evaluation can significantly enhance the effectiveness of assessment practices in English language teaching, particularly within vocational high school contexts.

Discussion

The findings of this study empirically demonstrate that systematic development and validation processes play a crucial role in improving the quality of English assessment instruments, particularly those designed to measure receptive skills. Assessment instruments that are developed without a clear framework often suffer from construct underrepresentation and weak measurement accuracy. By contrast, a structured development process beginning with syllabus analysis, followed by item design, expert validation, piloting, and item analysis ensures that test items function as intended. This finding is consistent with classical measurement theory, which emphasizes that validity and reliability are not inherent properties of a test but are achieved through rigorous design and evaluation procedures (Golzar et al., 2024; Sangsawang, 2015). Empirical studies on educational assessment have also shown that instruments subjected to iterative validation cycles demonstrate higher reliability coefficients and more balanced item difficulty levels, thereby producing more trustworthy test results. Thus, systematic development is not merely a technical requirement but a pedagogical necessity for ensuring accurate measurement of students' language competence.

Furthermore, aligning test items with syllabus indicators is essential for ensuring fairness and relevance in measuring students' receptive skills. Syllabus-based assessment reflects the principle of content validity, which requires that test items adequately represent the learning objectives and materials taught in the classroom (Rodríguez-Fuentes & Swatek, 2022). When assessment items are disconnected from syllabus indicators, students may be evaluated on competencies they were not explicitly taught, leading to biased or misleading interpretations of achievement. In the context of English language teaching, receptive skills such as reading and listening are closely tied to specific indicators, including social function, text structure, and linguistic features. Previous research has confirmed that alignment

* Corresponding author
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

between curriculum objectives and assessment instruments enhances transparency, accountability, and instructional coherence (Sato, 2022). Therefore, syllabus-aligned test item banks serve not only as evaluation tools but also as mechanisms for reinforcing curriculum implementation and instructional quality.

In addition, the inclusion of listening tasks addresses a critical gap in prior assessment practices and supports a more comprehensive evaluation of language proficiency. Many school-based English assessments tend to prioritize reading while neglecting listening, despite the fact that listening is a fundamental input skill in second language acquisition (Wardhana & Muhammad, 2021). According to Cummins (1979), listening provides the primary source of linguistic input that enables learners to develop vocabulary, grammar, and pragmatic awareness. Excluding listening components from assessment instruments results in an incomplete representation of students' receptive competence. Recent empirical studies have highlighted that balanced assessment of reading and listening yields a more accurate profile of learners' language ability and better informs instructional decision-making (Bennett, 2017). By integrating listening tasks into the test item bank, this study responds to contemporary assessment demands and aligns classroom evaluation practices with established theories of communicative language competence.

Other than that, the improvement in reliability coefficients from the first to the second trial provides strong empirical evidence that expert feedback and systematic revision cycles play a critical role in enhancing the quality of assessment instruments. In the first trial, the reliability index was categorized as moderate, indicating that although the test items functioned reasonably well, there were still inconsistencies in measuring students' receptive skills. After incorporating expert suggestions related to item construction, language clarity, distractor effectiveness, and alignment with learning indicators, the second trial demonstrated a substantial increase in reliability. This progression reflects the principle that reliability is not an inherent property of a test, but a characteristic that can be strengthened through iterative development and empirical refinement (Schulze, 2004). From an assessment theory perspective, repeated cycles of validation and revision enable test developers to reduce random measurement error and improve score consistency across administrations, thereby ensuring that test results more accurately reflect students' true language ability.

Furthermore, these findings reinforce the broader assessment literature which emphasizes that expert judgment is indispensable in test development, particularly in educational contexts where teachers often construct their own assessment tools. Cheong et al., (2023) argue that expert review contributes not only to content validity but also indirectly to reliability by eliminating ambiguous wording and poorly functioning items. In line with this view, the present study demonstrates that expert-driven revisions, followed by empirical trials, lead to measurable improvements in test performance. This confirms that reliability enhancement is a cumulative process grounded in both theoretical expertise and data-based decision making, rather than a one-time technical calculation.

In addition, the results of this study strongly support previous research highlighting the strategic importance of item banking and item analysis in improving educational assessment quality. Studies by Herwin & Nurhayati (2021) & Mustaqim et al., (2021) emphasize that well-managed item banks allow educators to store, monitor, and reuse test items with known psychometric properties, such as difficulty level and reliability indices. By systematically analyzing items after each administration, educators can identify weak or misleading questions and replace them with more effective alternatives. The current findings align with these studies, as the developed item bank demonstrated improved reliability and a more balanced distribution of item difficulty after analysis and revision, indicating higher

* Corresponding author

overall assessment quality.

Moreover, item banking supported by item analysis contributes to assessment fairness and curriculum alignment, which are central concerns in contemporary educational evaluation. According to Arikunto (2021), assessments that are not empirically analyzed risk misrepresenting students' actual competencies, leading to biased instructional decisions. The present study provides empirical confirmation that item banking, when combined with rigorous item analysis, enables teachers to design assessments that are consistent, objective, and aligned with syllabus indicators. This is particularly relevant in the context of English language assessment, where receptive skills require carefully constructed items to accurately measure comprehension processes rather than surface-level recall.

Overall, this study provides empirical evidence that a systematically developed test item bank can significantly enhance teachers' capacity to conduct efficient, objective, and curriculum-based evaluations, particularly in the context of English language assessment. The findings of this study align with this view, as the developed item bank demonstrated acceptable reliability indices and balanced difficulty levels, indicating that it can produce stable and trustworthy measurement results. Yet, the alignment between item banks and curriculum requirements strengthens the validity of classroom assessment. Curriculum-based evaluation requires that test items reflect the stated competencies, learning indicators, and language skills outlined in the syllabus. According to Sun & Park (2023), a test can be considered valid in terms of content if it represents a comprehensive sample of the instructional objectives being taught. The item bank developed in this study was explicitly designed by mapping each test item to specific curriculum indicators, ensuring that both reading and listening skills were adequately assessed. This systematic alignment reduces the risk of construct underrepresentation and ensures that evaluation outcomes are pedagogically meaningful for both teachers and students.

Toba et al., (2019) argue that item banks enable the development of parallel tests, which are essential for monitoring learning progress over time and maintaining consistency across different test administrations. In this study, the improvement in reliability scores between the first and second trials illustrates how iterative item analysis and revision contribute to higher-quality assessment instruments. Overall, the findings of this study confirm that well-developed item banks function as an effective assessment infrastructure in educational practice. They enhance efficiency by reducing teachers' workload in test construction, promote objectivity through empirically validated items, and ensure curriculum alignment by systematically integrating learning objectives into assessment design. Grounded in established theories of educational measurement and supported by empirical evidence, this study reinforces the argument that item bank development should be institutionalized as a sustainable strategy for improving the quality of classroom evaluation.

CONCLUSIONS AND SUGGESTION

The present study empirically demonstrates that a receptive-skills test item bank can be systematically developed to meet the requirements of validity, reliability, and syllabus alignment for vocational high school English assessment. Through a structured research and development approach, the test items were constructed based on competency standards, indicators, and learning objectives outlined in the national curriculum. Expert validation and iterative revisions ensured strong content validity, confirming that the items accurately represented the targeted reading and listening competencies. This finding supports established assessment theory, which emphasizes alignment between instructional objectives

*** Corresponding author**
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

and assessment instruments as a prerequisite for meaningful evaluation.

The developed instrument effectively measures students' receptive skills by integrating reading and listening components in a balanced and standardized format. Such integration reflects theoretical perspectives that position receptive skills as foundational inputs for language acquisition. Empirical analysis further indicated acceptable to high reliability indices, demonstrating consistency of measurement across test administrations. Additionally, the distribution of item difficulty levels suggests that the test items are neither excessively easy nor overly difficult, aligning with classical test theory principles. These characteristics confirm the instrument's capacity to generate accurate and fair representations of students' English proficiency.

Based on these outcomes, the test item bank is considered suitable for use in semester examinations at SMK N 5 Pekanbaru. Its structured design supports objective scoring, comparability of results, and efficient test construction. To sustain assessment quality, it is recommended that teachers routinely conduct item analysis before reusing test items, as continuous evaluation strengthens test accuracy and fairness. At the institutional level, schools should formalize item bank development as part of their assessment system to ensure long-term consistency and quality control. Finally, future research is encouraged to expand the item bank to include productive skills such as speaking and writing, and involve larger samples to enhance generalizability and contribute to more comprehensive language assessment practices.

REFERENCES

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi Aksara.

Bennett, M. J. (2017). Developmental Model of Intercultural Sensitivity. In *The International Encyclopedia of Intercultural Communication* (pp. 1–10). https://doi.org/https://doi.org/10.1002/9781118783665.ieicc0182

Bhoki, M. I. (2025). Peran Kepala Sekolah dalam Mengoptimalkan Kemampuan Guru untuk Mengembangkan Tujuan Pembelajaran Berdasarkan Kebutuhan Siswa di SMPN 2 Golewa kolektif melalui strategi kepemimpinan yang efektif . Kepala sekolah di SMPN 2 Golewa ,. *Jurnal Sadewa: Publikasi Ilmu Pendidikan, Pembelajaran Dan Ilmu Sosial*, *3*(1), 211–240. https://doi.org/10.61132/sadewa.v3i1.1511

Byram, M., Holmes, P., & Savvides, N. (2013). Intercultural Communicative Competence in Foreign Language Education: Questions of Theory, Practice and Research. *Language Learning Journal*, *41*(3), 251–253. https://doi.org/10.1080/09571736.2013.836343

Cheong, H., Lyons, A., Houghton, R., & Majumdar, A. (2023). Secondary Qualitative Research Methodology Using Online Data within the Context of Social Sciences. *International Journal of Qualitative Methods*, *22*, 16094069231180160. https://doi.org/10.1177/16094069231180160

Chung, E., & Wan, A. (2025). Examining the use of academic vocabulary in first-year ESL undergraduates' writing: A corpus-driven study in Hong Kong. *Assessing Writing*, *63*, 100913. https://doi.org/https://doi.org/10.1016/j.asw.2024.100913

Cummins, James. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

*** Corresponding author**

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction*. Pearson/Allyn & Bacon. https://books.google.co.id/books?id=l9JfQgAACAAJ

Golzar, J., Yacoub, O., & McKinley, J. (2024). E-learning successes with English language teachers in under-resourced non-WEIRD contexts. *International Journal of Applied Linguistics*, *34*(3), 1159–1182. https://doi.org/https://doi.org/10.1111/ijal.12557

Hartell, E., & Buckley, J. (2021). *Comparative Judgment: An Overview BT - Handbook for Online Learning Contexts: Digital, Mobile and Open: Policy and Practice* (A. Marcus-Quinn & T. Hourigan (eds.); pp. 289–307). Springer International Publishing. https://doi.org/10.1007/978-3-030-67349-9_20

Herwin, & Nurhayati, R. (2021). Measuring students' curiosity character using confirmatory factor analysis. *European Journal of Educational Research*, *10*(2), 773–783. https://doi.org/10.12973/EU-JER.10.2.773

Hollister, B., Nair, P., Hill-Lindsay, S., & Chukoskie, L. (2022). Engagement in Online Learning: Student Attitudes and Behavior During COVID-19. *Frontiers in Education*, *7*(May). https://doi.org/10.3389/feduc.2022.851019

Hutchins, T. L., Knox, S. E., & Fletcher, E. C. (2024). Natural language acquisition and gestalt language processing: A critical analysis of their application to autism and speech language therapy(). *Autism & Developmental Language Impairments*, *9*, 23969415241249944. https://doi.org/10.1177/23969415241249944

Leung, C. H., & Cheng, S. C. L. (2019). An Empirical Study on Integration of Experiential Learning and Mobile Learning. *Asian Journal of Empirical Research*, *9*(4), 88–98. https://doi.org/10.18488/journal.1007/2019.9.4/1007.4.88.98

Liu, C., & Chen, M. (2022). A genre-based approach in the secondary school English writing class: Voices from student-teachers in the teaching practicum. *Frontiers in Psychology*, *Volume 13-2022*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.992360

Mustaqim, M. A., Almarzuqi, M. F., & Sibilana, A. R. (2021). *Education in Psychomotoric Aspect and Creative Development BT - Proceedings of the International Conference on Engineering, Technology and Social Science (ICONETOS 2020)*. 636–645. https://doi.org/10.2991/assehr.k.210421.093

Rodríguez-Fuentes, R. A., & Swatek, A. M. (2022). Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students' grammatical construction learning. *System*, *104*, 102676. https://doi.org/https://doi.org/10.1016/j.system.2021.102676

Roebianto, A., Savitri, I., Sriyanto, A., Syaiful, I., & Mubarokah, L. (2023). Content Validity: Definition and Procedure of Content Validation in Psychological Research. *TPM - Testing*, *30*, 5–18. https://doi.org/10.4473/TPM30.1.1

Romadin, A., Nuhadi, D., & Yoto, Y. (2022). Implementation of Work Based Learning on Welding Engineering Expertise Competency in The Manufacturing Industry. *Journal of Vocational Education Studies*, *5*(1), 16–31. https://doi.org/10.12928/joves.v5i1.5674

Sangsawang, T. (2015). Instructional Design Framework for Educational Media. *Procedia - Social and Behavioral Sciences*, *176*, 65–80. https://doi.org/https://doi.org/10.1016/j.sbspro.2015.01.445

*** Corresponding author**
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx

Sato, T. (2022). Assessing Critical Thinking through L2 Argumentative Essays: An Investigation of Relevant and Salient Criteria from Raters' Perspectives. *Language Testing in Asia*, *12*(1), 9. https://doi.org/10.1186/s40468-022-00159-4

Schulze, R. (2004). Meta-analysis: A comparison of approaches. In *Meta-analysis: A comparison of approaches.* Hogrefe & Huber Publishers.

Shoda, V. P., & Yamanaka, T. (2022). A Study on Instructional Humor: How Much Humor Is Used in Presentations? In *Behavioral Sciences* (Vol. 12, Issue 1). https://doi.org/10.3390/bs12010007

Siregar, L. K., Mayuni, I., & Rahmawati, Y. (2023). Culturally responsive English teaching: Developing a model for primary school EFL teachers in Indonesia. *Issues in Educational Research*, *33*(4), 1582–1600.

Sugiyono, S. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Sun, W., & Park, E. (2023). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. In *Sustainability* (Vol. 15, Issue 17). https://doi.org/10.3390/su151713242

Taye, T., & Mengesha, M. (2024). Identifying and Analyzing Common English Writing Challenges among Regular Undergraduate Students. *Heliyon*, *10*(17). https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e36876

Toba, R., Noor, W. N., & Sanu, L. O. (2019). The Current Issues of Indonesian EFL Students' Writing Skills: Ability, Problem, and Reason in Writing Comparison and Contrast Essay. *Dinamika Ilmu*, *19*(1), 57–73. https://doi.org/10.21093/di.v19i1.1506

Wardhana, Y. Y., & Muhammad, R. N. (2021). An Investigation of Video and Audio to Improve Students' Motivation in Learning Listening during Online Learning at Ban Pongneeb School, Thailand. *Journal of English Teaching, Literature, and Applied Linguistics*, *5*(1), 20–27.

Widiyanti, R. (2024). Evaluation of the Implementation of Field Work Practices Using the CIPP Model. *Sinergi International Journal of Education*, *2*(1), 1–11. https://doi.org/10.61194/education.v2i1.122

Widyastuti, R., & Utami, I. S. (2018). Development of Product-Based Job Sheet as Instructional Media in Vocational Education. *Journal of Educational Science and Technology (EST); Volume 4 Number 2 August 2018DO - 10.26858/Est.V4i2.6064* . https://ojs.unm.ac.id/JEST/article/view/6064

Zhang, T., & Zhang, L. J. (2021). Taking Stock of a Genre-Based Pedagogy: Sustaining the Development of EFL Students' Knowledge of the Elements in Argumentation and Writing Improvement. In *Sustainability* (Vol. 13, Issue 21, p. 11616). https://doi.org/10.3390/su132111616

*** Corresponding author**
IJPER (International Journal of Pedagogy and Education Research), x (x), xxxx, xx-xx
P-ISSN: XXXX-XXXX, E-ISSN: XXXX-XXXX | DOI: http://doi.org/xx.xxxxx/ijper.vxix.xxxx