



WORKPLACE SIMULATION-BASED ENGLISH FOR SPECIFIC PURPOSES (ESP) ASSESSMENT IN VOCATIONAL EDUCATION: A SCOPING REVIEW

Rindilla Antika*, Doris Sukma
Politeknik Negeri Bengkalis, Riau
rindillaantika@polbeng.ac.id*

Abstract

English for Specific Purposes (ESP) plays an important role in workplace competence in vocational education. Work simulation is widely promoted as a way to provide authentic, performance-based assessment. However, it remains unclear whether simulation is actually used to assess ESP proficiency or mainly to support learning, and whether the assessment instruments are properly validated. This study uses a scoping review based on the PRISMA-ScR guidelines to examine how work simulation is used in ESP assessment in vocational education. A systematic search of four databases (Lens.org, ERIC, DOAJ, and Garuda) for studies published between 2016 and 2026 identified 172 records. After removing duplicates and conducting two-stage screening, 11 studies met the inclusion criteria. The findings show that work simulation is used mainly as a learning strategy rather than as a way to assess competence. Only one of the 11 studies conducted assessment within a simulated task and validated its measurement instrument. Other studies often treated statistical significance as evidence of validation, validated materials or satisfaction questionnaires instead of proficiency instruments, or used standardized tests without providing a context-specific validity argument. These findings highlight a gap between the authenticity promised by simulation-based assessment and its actual use in ESP. Future research should prioritize developing and validating assessment instruments embedded in simulated work tasks, including emerging AI- and virtual-reality-based environments.

Keywords: Assessments ESP, Instruments Validation, Scoping Review, Vocational Education, Work-Simulation.

INTRODUCTION

Vocational and technical education serves a distinct purpose compared with general education. It is not only concerned with providing theoretical knowledge but also with developing graduates who are able to apply their competencies in real workplace settings (Nurjanah & Ana, 2022; OECD, 2021). In this context, professional competence increasingly involves the ability to communicate effectively in English. However, the required proficiency is not limited to general English skills; it involves English for Specific Purposes (ESP), which is developed according to the communication demands of particular occupations. ESP connects language learning with workplace practices by preparing learners to use English appropriately within their professional environments (Coxhead et al., 2019; Koester, 2021). In vocational fields such as maritime studies, hospitality, nursing, and business, English proficiency has become an essential competency that supports workplace

safety, service effectiveness, professional interaction, and career development. Therefore, ESP is increasingly recognized as an integral component of vocational education in preparing graduates to meet industry expectations.

Unlike general English, which focuses on broad communication skills, ESP is developed based on language needs in specific academic or professional fields. The main difference in ESP lies not only in the use of technical vocabulary but also in its focus on communicative actions in specific work situations (Saputro & Savitri, 2025; Xenia & Indriyani, 2025). The success of ESP learners is not solely determined by their ability to construct sentences correctly, but rather by their ability to use language to complete occupational communication tasks, such as greeting customers, reporting on a patient's condition, conducting business negotiations, or explaining workplace safety procedures. Thus, ESP learning objectives are performative in nature because they emphasize the demonstration of language proficiency in real-world contexts, rather than merely the mastery of isolated linguistic elements (Hidayati & Meisani, 2023). This shift in orientation has significant implications for ESP assessment, as communicative competence in a work context requires a different assessment approach than that used for general language proficiency.

The consequences of this performative orientation of ESP are evident in the assessment aspect. Language proficiency that is ready for use in a work context cannot be fully represented through written tests that are detached from the actual language-use situation. Such tests can provide insight into learners' knowledge of linguistic elements, but they often fail to demonstrate their ability to use language to complete communication tasks under the demands of specific work situations (Knoch & Macqueen, 2016; Liao et al., 2023). This situation has driven a shift toward authentic and performance-based assessment—an approach that places learners in tasks reflecting real-world communication practices and job demands (Dahlback et al., 2020; Norris et al., 2023). It is within this framework that work simulations are increasingly being considered as an approach to create language use contexts that more closely resemble the professional world.

Work simulations—role-playing, the use of simulators, and scenario-based tasks—promise to bridge the gap between the classroom and the workplace. They present occupational communication demands in a controlled yet realistic environment, so that skills can, in principle, be assessed under conditions that closely resemble reality without the full consequences of the actual workplace (Jossberger et al., 2017; Seifried et al., 2020). Because of its closeness to authentic tasks, simulation is often advocated as a vehicle for authentic ESP assessment—that is, assessment that reflects the real communicative demands of a profession (Dahlback et al., 2020; Sari et al., 2024). Such an approach appears particularly relevant for vocational education, where competence is expected to be demonstrated through the ability to perform tasks within specific professional contexts. However, a promise is not proof. The fact that simulations provide an authentic learning context does not in itself mean that they are actually used—let alone validated—as a means of measuring proficiency.

It is at this stage that conceptual issues in the literature begin to emerge. Terms such as simulation, role-play, authentic assessment, and workplace-based assessment are often used interchangeably, even though each has distinct characteristics and functions in both the learning and assessment processes (Anderson et al., 2021; Rietmeijer et al., 2025). As a result, it is not always easy to distinguish studies that use simulation as a means to obtain evidence of competence from those that merely use it as a learning strategy, while the process of measuring ability is still conducted through conventional instruments. Another issue arises when a study states that competence is assessed through simulation but fails to

adequately explain how the assessment instrument was developed, used, and validated. Without clarity regarding the role of simulation in the assessment process and the quality of the instruments used, it remains difficult to verify the validity of claims regarding authentic assessment.

These conceptual ambiguities are accompanied by a more practical problem: the existing evidence remains fragmented across different educational contexts. Studies on simulation-based ESP assessment have appeared at various levels of vocational education, from vocational secondary schools to polytechnics and other forms of vocational higher education (Mukadimah & Sahayu, 2021; Putra et al., 2021; Sumira et al., 2025), as well as in countries with different vocational education traditions. Although simulation is frequently proposed as a way to support authentic assessment, the available literature has not yet been brought together to examine how these assessments are designed, whether the instruments used have been validated by their authors, and which areas of research remain underexplored. Previous reviews have tended to focus on related approaches, including virtual reality, mobile learning, and artificial intelligence in language education, rather than examining simulation specifically as a means of assessing ESP (Crompton et al., 2024; Rafiq et al., 2021; Rahmanu & Molnár, 2026). Given the emerging and heterogeneous nature of this evidence, a scoping review is appropriate because it allows the existing literature to be mapped systematically without reducing the inquiry to a narrowly defined question of intervention effects (Arksey & O'Malley, 2005; Munn et al., 2018; Tricco et al., 2018).

This gap is particularly relevant in the context of Indonesian vocational education. Vocational schools and polytechnic institutions in Indonesia have increasingly focused on improving graduates' work readiness and ensuring that their competencies correspond with industry expectations (Suharno et al., 2020). In line with this direction, simulation-based learning approaches have been increasingly introduced as a means of providing students with experiences that resemble professional practices (Pradipta et al., 2021; Prianto et al., 2021). However, the development of appropriate assessment practices requires a clearer understanding of how simulation has been utilized to generate evidence of competence and how the validity of assessment instruments has been established. Without such evidence, attempts to develop authentic ESP assessment may rely more on assumptions about simulation than on systematically identified practices. Therefore, this scoping review seeks to map the existing evidence on work simulation as an ESP assessment approach in vocational education, with particular attention to assessment designs, implementation practices, instrument validation processes, and remaining gaps in the literature. By adopting a scoping review approach, this study aims to provide a comprehensive overview of how simulation-based ESP assessments have been conceptualized and applied across vocational education contexts. Rather than evaluating the effectiveness of simulation as either a teaching strategy or an assessment method, this review examines how these two functions have been distinguished and represented within the existing body of research.

METHOD

Design and Framework

This study employed a scoping review design, following the framework developed by Arksey and O'Malley (2005) and refined by Levac et al. (2010). The review was conducted in accordance with the Joanna Briggs Institute (JBI) guidelines for scoping reviews (Peters et al., 2021), while the reporting process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR)

guidelines (Tricco et al., 2018). This design was chosen because the research objective was to map the characteristics and scope of the available evidence, rather than to test the effectiveness of an intervention or calculate the magnitude of an effect. A review protocol was developed to guide the review process. The protocol specified the research questions, PCC framework, search strategy, and data-charting variables. The protocol was not registered in a public registry. PROSPERO primarily accepts protocols for systematic reviews addressing health-related outcomes and does not currently register scoping review protocols (Centre for Reviews and Dissemination, 2025). Scoping review protocols may also be registered on platforms such as the Open Science Framework (OSF). In the absence of public registration, the protocol is available from the corresponding author upon reasonable request. The literature on simulation-based ESP assessment is characterized by conceptual and terminological heterogeneity, in which terms such as simulation, role-play, authentic assessment, and workplace-based assessment are used in various contexts with different emphases. Therefore, an initial mapping is needed to identify patterns in assessment design, the characteristics of the instruments used, and gaps requiring further investigation.

This study is guided by three research questions focusing on: (1) how simulation-based ESP assessment designs and methods are applied in vocational education; (2) how the validation process of assessment instruments is reported in previous studies; and (3) what research gaps still exist in the available literature. The second research question was formulated to examine the existence and nature of the instrument validation process, rather than to conduct an in-depth assessment of the quality of validity. This approach was adopted because reporting practices concerning instrument development and validation vary across studies, necessitating a systematic identification of these practices before conducting further methodological evaluation.

Eligibility Criteria

The eligibility criteria were developed based on the Population–Concept–Context (PCC) framework in accordance with the JBI scoping review guidelines (Peters et al., 2021). The Population includes vocational education learners, ranging from vocational high schools to polytechnics. The “Concept” focuses on English for Specific Purposes (ESP) assessments involving work simulations as well as information related to the development or validation of assessment instruments. The “Context” encompasses the vocational education environment or technical and vocational education and training (TVET). In line with the objectives of a scoping review, methodological quality appraisal was not used as a basis for study inclusion or exclusion (Peters et al., 2021). However, data integrity issues—such as inconsistencies in information, discrepancies in data counts, or problems with source identification—were still noted during the data extraction process to ensure the transparency of the review. Each study excluded during the full-text review stage was assigned a single reason for exclusion in accordance with the PRISMA-ScR guidelines. Details of the criteria are presented in Table 1.

Table 1: Inclusion Criteria and Exclusion Codes

Code	Inclusion Criteria	Reason for Exclusion
I1	Studies in the context of vocational education/TVET	E1: studies outside the vocational context
I2	ESP / vocational domain-specific English, not general English	E2: not ESP; general English

I3	A work simulation component as an intervention or assessment	E3: no simulation component
I4	A language proficiency assessment that is administered and measured—not self-reported perception	E4: no assessment component
I5	Empirical study, published 2016–2026	E5: not empirical / outside the date range
—	Full text is accessible	Full text is not available

Two main criteria were applied in the study selection process. First, simulations must be used as part of language proficiency assessment, not merely as a learning activity or curriculum element. Studies that only measured participants' perceptions or mentioned simulations without implementing an assessment were excluded. Second, full-text access was a requirement for inclusion, as selection decisions were not based on abstracts alone. Studies without full-text access were excluded on the grounds that "full text was not available."

Information Sources and Search Strategy

The literature search was conducted using four databases: Lens.org, ERIC, DOAJ, and Garuda. Lens.org was used for interdisciplinary coverage, ERIC for educational research, DOAJ for open-access publications, and Garuda to capture Indonesian studies that may not be indexed internationally. The search strategy combined three conceptual groups: (1) English for Specific Purposes (ESP), (2) vocational education (e.g., TVET, SMK, and polytechnics), and (3) work simulation and context-based assessment (e.g., work simulation, role-play, simulation-based learning, and authentic assessment). Boolean operators were applied with a publication range of 2016–2026, and the complete search strings are provided in Appendix 1. Technical limitations were identified, including ERIC's restricted year-filtering function, which limited retrieval to 2017–2026, and Garuda's lack of citation export features, requiring manual metadata collection. These limitations were documented as part of the retrieval process and were not used as exclusion criteria.

Study Selection, Data Charting, and Evidence Synthesis

All records identified through the literature search were managed using Rayyan for deduplication and eligibility screening. Following the removal of 12 duplicate records, study selection was conducted in two stages: title and abstract screening, followed by full-text assessment against the predefined inclusion and exclusion criteria. A primary reason for exclusion was recorded for each study excluded at the full-text stage. Record matching was verified using article titles and author names to minimize identification errors arising from variations in file naming and reference formatting.

Data were extracted using a structured form covering study characteristics, including publication source, educational level, country, simulation type, assessment procedures and instruments, validity evidence, validation status, study design, key findings, and reporting quality. The extracted data were synthesized using narrative and descriptive approaches. For RQ2, instrument validation was classified as Yes, Yes (weak), Partially, or No, based on the validity or reliability evidence reported by the study authors. The use of externally developed standardized instruments or statistical analyses alone was not considered sufficient evidence of instrument validity unless supporting evidence was explicitly reported in the study.

For RQ1 and RQ3, simulation use was classified into three categories: assessment-based simulations, in which the primary purpose was to evaluate learners' performance; assessment-contextual simulations, in which simulation provided the context or environment for assessment; and purely pedagogical simulations, in which simulation was employed primarily as an instructional strategy rather than as an assessment tool. This classification distinguished simulations that directly generated evidence of learners' ESP competence from those used primarily for instructional purposes. Accordingly, the synthesis examined not only whether simulations were incorporated but also their function within the assessment process and the extent to which the instruments used to measure competence were supported by validity and reliability evidence..

PRISMA-ScR Flowchart

The study selection process followed the PRISMA-ScR flowchart. The initial search yielded 172 records from Lens.org (n = 115), ERIC (n = 44), DOAJ (n = 1), and Garuda (n = 12). After removing 12 duplicates, 160 records underwent title and abstract screening. At this stage, 83 records were excluded because they did not align with the study's focus. A total of 77 articles were then evaluated through full-text review. Of these, 66 articles were excluded for failing to meet the inclusion criteria: lack of a work simulation component (n = 32), no relevance to ESP (n = 14), absence of an assessment component (n = 8), unavailability of the full text (n = 7), and failure to meet design criteria or publication year ranges (n = 5). A total of 11 studies ultimately met all criteria and were included in the final synthesis. Details of the selection process are presented in Figure 1.

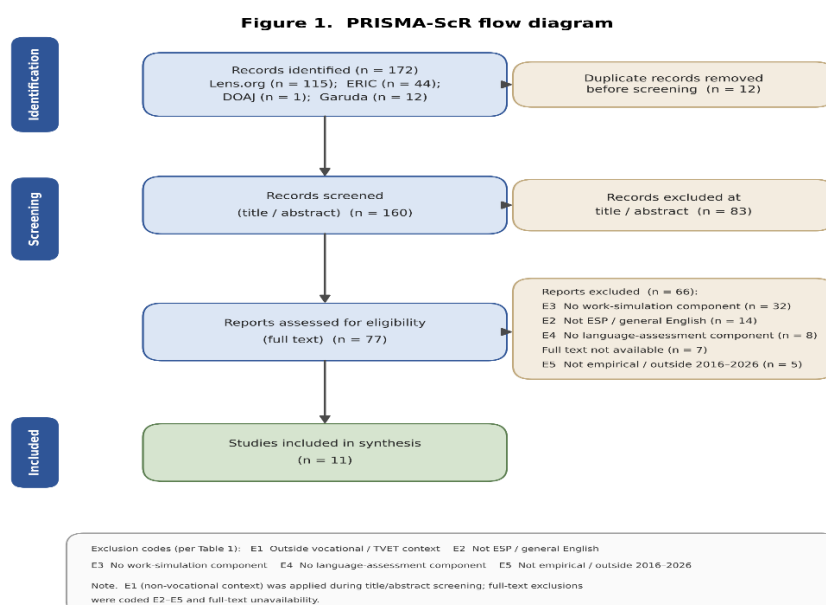


Figure 1. PRISMA-ScR Flowchart

RESULTS

A total of eleven studies met all inclusion criteria and were analyzed through a narrative-descriptive synthesis based on three predefined research questions. The distribution of these studies' characteristics is presented in Table 2, grouped according to the role of simulation in the assessment process, ranging from simulations used as a basis for measuring competencies to those functioning primarily as learning strategies. This categorization reveals an important pattern in the reviewed literature. Although simulations are often

viewed as having great potential to support authentic ESP assessment, research that explicitly positions simulations as the primary mechanism for generating evidence of competence remains relatively limited.

Table 2: Matrix of included studies (n=11)

Code	Study	Context (Level, Country)	Design	Role of simulation	Language Proficiency Instrument Validation
1	Huang (2026)	Vocational Higher School - China	Controlled experiments + action research (mixed)	Assessment- based	Yes
2	Kanca et al. (2021)	Polytechnic - Indonesia	CAR + quasi experiment (mixed)	Assessment- based	Yes (weak)
3	Su et al. (2022)	Nursing University- Taiwan	Action research (mixed)	Assessment- contextual	Partially
4	Lo (2017)	Vocational School - Taiwan	Qualitative case study	Assessment- contextual	No (formal)
5	Guo & Yu (2022)	Health Vocational School - China	Quasi-experiment (mixed)	Assessment- contextual	No (external)
6	Zinchenko et al. (2020)	Technical University - Ukraine	Pre-post experiment	Assessment- contextual	No
7	Tyas et al. (2020)	Vocational Higher School - Indonesia	pre-experiment one-group (mixed)	Assessment- contextual	No
8	Rattanaovich (2017)	Vocational School - Thailand	Quasi-experiment	Assessment- contextual	No (external)
9	Abdullah & Soliman (2019)	Vocational School - Egypt	Quasi-experiment	Assessment- contextual	No
10	Ramendra et al. (2025)	Tourism Vocational School - Indonesia	R&D / ADDIE	Purely pedagogical	No

11	Apriyanti et al. (2021)	Polytechnic Indonesia	-	R&D (Gall Borg)	&	Purely pedagogical	No
----	-------------------------	-----------------------	---	-----------------	---	--------------------	----

Characteristics of ESP Assessment Studies Based on Work Simulations

The design patterns and assessment methods for ESP based on work simulations indicate that most studies still focus on the perspective of learning effectiveness rather than on the development and validation of assessments. The majority of studies use an intervention design to test changes in students’ abilities after the implementation of simulations, either through a quasi-experimental approach with a pre–post design or through development research such as Research and Development (R&D) and Design-Based Research (DBR). Some studies using quasi-experimental designs focus on measuring score improvements after the simulation intervention are implemented (Abdullah & Soliman, 2019; Guo & Yu, 2022; Huang, 2026; Tyas et al., 2020; Zinchenko et al., 2020), while development studies place greater emphasis on the process of designing simulation-based learning models or media (Apriyanti et al., 2021; Ramendra et al., 2025). Only a small fraction of studies has used qualitative approaches to understand students’ experiences or perceptions regarding the use of simulations (Lo, 2017). The predominance of these intervention designs suggests that simulations are more often treated as pedagogical interventions whose impact on learning outcomes needs to be tested, while the tools used to demonstrate competency have not yet become a primary focus of research.

In terms of implementation, simulations in ESP instruction and assessment are primarily carried out through activities that mimic real-world professional interactions. Role-play, business simulations, and work-situation-based dialogs are the most frequently used forms in various studies (Guo & Yu, 2022; Huang, 2026; Kanca et al., 2021; Tyas et al., 2020; Zinchenko et al., 2020). In addition, several studies have developed task-based approaches and authentic tasks designed to approximate communication practices in the workplace (Abdullah & Soliman, 2019; Apriyanti et al., 2021; Lo, 2017; Rattanavich, 2017; Su et al., 2022). Technological advancements are also beginning to influence simulation design, particularly through the use of artificial intelligence and virtual environments. However, AI-based and virtual simulation approaches are still in the early stages of development, as seen in Huang's (2026) study, which utilizes generative AI models, and the research by Ramendra et al. (2025), which developed an AI-based virtual tour. These findings indicate a new direction in the development of ESP simulations, although it cannot yet be considered a dominant practice in the current literature.

The most important aspect in understanding assessment patterns is how simulations are structured within research. Most studies use simulations as part of the learning process, while the use of simulations as the primary source of assessment evidence remains relatively rare. Huang (2026) is one of the studies that explicitly uses students’ performance in a simulation environment as the basis for assessment through situation-based oral communication tasks, employing an assessment instrument that considers multiple dimensions of performance. The study by Kanca et al. (2021) also demonstrates a similar trend through the use of scenario-based oral tests in the field of Room Division. However, some other studies use simulations only as a context or stimulus in assessment tasks, while the scoring process is still conducted using conventional instruments (Guo & Yu, 2022; Lo, 2017; Su et al., 2022). In another group of studies, simulation’s function more as a pedagogical strategy to enhance the learning experience, while assessment is still conducted through separate methods such as pre- and post-test assessments (Abdullah & Soliman, 2019; Apriyanti et al., 2021; Ramendra et al., 2025; Rattanavich, 2017). Thus, the existing

literature indicates that work simulations are more often developed as a teaching method rather than as a means of directly assessing ESP proficiency.

Mapping the Validation of Work-Simulation-Based ESP Assessment Instruments

From the perspective of instrument validation, most of the studies analyzed did not demonstrate an adequate validation process for the language proficiency assessment tools used. Of the eleven studies reviewed, only Huang's (2026) study explicitly tested the validity and reliability of the developed language proficiency assessment instrument, yielding a Cronbach's α reliability coefficient of 0.872 and a cumulative variance of 76.35% from the Exploratory Factor Analysis (EFA) (Huang, 2026). Meanwhile, other studies tended to use assessment instruments without providing empirical evidence regarding their measurement quality. This situation indicates that research attention is still more focused on the implementation of simulations or the impact of learning, rather than on the development of instruments capable of producing valid evidence regarding students' ESP competencies.

Further analysis shows that weaknesses in instrument validation in these studies emerge in several patterns. First, there is a tendency to equate testing for differences in scores with the instrument validation process. Some studies report the use of t-tests, Student-Fisher criteria, or analysis of covariance (ANCOVA) to demonstrate the success of an intervention; however, these tests only indicate the statistical significance of changes or differences in scores, not the quality of the instrument in measuring the targeted language ability construct (Abdullah & Soliman, 2019; Ramendra et al., 2025; Tyas et al., 2020; Zinchenko et al., 2020). In the context of assessment, an instrument's validity is not determined solely by the presence of statistical differences, but by the extent to which the available evidence supports the interpretation of scores as a representation of the ability being measured (Kane, 2013).

Second, some studies validated objects that were not aligned with the purpose of measuring language proficiency. For example, the validations conducted by Ramendra et al. (2025) and Apriyanti et al. (2021) focused on the suitability of learning media or materials, rather than on ESP proficiency assessment instruments. A similar pattern was found in the studies by Su et al. (2022) and Kanca et al. (2021), which tested the validity and reliability of instruments measuring learner perception or satisfaction—such as the Customer Satisfaction Questionnaire (CSQ)—but these instruments were not used to measure language proficiency directly. The distinction between learning evaluation instruments and competency measurement instruments needs to be emphasized because they serve different measurement purposes.

Third, some studies use externally developed standardized tests and report their psychometric characteristics but do not validate the use of these instruments in the context of their research. For example, Guo & Yu (2022) used the METS-1, while Rattanavich (2017) used an institutional reading test as a measure of language proficiency. Although these instruments already have reliability data from their original developers, their use in different contexts, populations, and for different assessment purposes still requires its own validity justification. Thus, the existence of a standardized instrument does not automatically mean that it has been validated for measurement purposes in a specific study.

Research Gaps and the Development for Work-Simulation-Based ESP Assessment

The evidence review reveals several significant gaps in research on simulation-based ESP assessment. First, studies that truly use simulations as the basis for assessment remain very limited. Most research uses simulations as a learning approach, rather than as a tool to generate measurable evidence of competence. This situation highlights the need to develop

assessments specifically designed for work simulation contexts and supported by adequate validity evidence. Second, limitations in instrument validation indicate that future research must place greater emphasis on developing measurement tools capable of capturing professional communication skills, rather than relying solely on standardized tests or instruments not adapted to the simulation context. Third, existing research still focuses more on the effectiveness of learning methods than on the quality of the assessments themselves. Furthermore, research evidence remains concentrated in certain regions of Asia, particularly East Asia and Southeast Asia; therefore, the application of research findings to different vocational education contexts must be approached with caution. Advances in technology-based simulations, such as artificial intelligence (AI) and virtual reality (VR), also open up new opportunities, but the validity of the instruments used still requires further study (Huang, 2026; Ramendra et al., 2025).

In the mapping process, several studies also highlight points that require attention in the interpretation of evidence. The study by Abdullah & Soliman (2019), for example, contains inconsistencies between the results table and the interpretation of the data; therefore, its contribution to claims of effectiveness should be treated with caution. Tyas et al. (2020) also noted discrepancies between the description of the TEVOCS instrument and the characteristics of the study participants. Additionally, there were several reporting limitations in the study by Ramendra et al. (2025), such as incomplete table information, as well as in Lo (2017), which referenced the psychometric characteristics of the instrument from a companion study. Within the framework of a scoping review, these observations were not used as grounds for excluding studies but were considered when assessing the strength of the available evidence and highlight the importance of more transparent reporting on instrument validation in simulation-based ESP assessment research.

DISCUSSIONS

This review maps out how work-based ESP assessments using simulations are designed, implemented, and supported by instrument validation evidence. Key findings indicate that in vocational education, simulations are more commonly used as a learning approach rather than as an assessment method that systematically generates evidence of competency. Of the eleven studies analyzed, only one used simulation as the basis for measuring competencies while also validating its assessment instrument. As a scoping review, this study does not aim to evaluate the effectiveness of a particular method but rather to describe patterns of simulation use, identify gaps in the available evidence, and suggest directions for future research. In the context of ESP assessment, these findings underscore the need to make a conceptual distinction between simulation as a learning strategy and simulation as an assessment instrument that requires evidence of validity.

Instrument Validation Gaps

Only one of the eleven studies reviewed explicitly described the procedures used to evaluate the validity of its assessment instrument. This finding highlights limitations in methodological reporting and the limited availability of documented validity evidence for the instruments employed. The reviewed studies reveal recurring limitations in the reporting of validation procedures and the documentation of supporting evidence as a process of gathering evidence to support score interpretations and uses. Instrument validation requires evidence supporting the intended interpretations and uses of assessment scores, rather than reliance on statistical indicators of reliability alone. Contemporary validity theory conceptualizes validity in terms of the strength of the evidence supporting the interpretations

and uses of assessment scores, rather than as an inherent property of an instrument or an outcome established through a single statistical procedure (Kane, 2013; Newton & Shaw, 2014). This issue is particularly relevant to simulation-based assessment in English for Specific Purposes (ESP), where assessment scores must be supported by evidence that learners can communicate effectively and perform relevant professional tasks in simulated workplace settings. Without sufficient validity evidence, the extent to which these assessments measure the intended job-related competencies remains uncertain. Accordingly, claims regarding the effectiveness of simulation-based assessments in evaluating occupational competence should be substantiated by robust empirical evidence of validity and effectiveness rather than relying solely on assumptions derived from instructional principles

Simulation as an Intervention, Not an Assessment Tool

The limited attention given to instrument validation can be understood in relation to the dominant design orientation of the existing corpus. Most studies primarily focus on the effectiveness of instructional interventions, particularly whether a particular method improves learners' scores. Within this orientation, assessment instruments are often treated merely as tools for measuring change rather than as developed measures that require systematic validation. However, validity cannot be regarded as a consequence of statistical analysis alone. Rather, it concerns the extent to which available evidence and theoretical justification support the intended interpretation and use of assessment scores (Kane, 2013; Messick, 1989). Accordingly, the validation gaps identified in this review should not be viewed simply as technical omissions. Instead, they reflect a broader research orientation in which simulation is predominantly conceptualized as a learning intervention rather than as an assessment approach that requires explicit attention to the validity of its measurement processes.

The problem, therefore, lies not necessarily in the research design itself, but in the object and focus of validation. Although development-oriented approaches, including research and development (R&D) and design-based research, provide opportunities to develop new instruments, validation efforts in such studies are often directed primarily toward instructional media or learning materials rather than toward instruments intended to measure learner proficiency (Plomp & Nieveen, 2013). In the context of simulation-based ESP assessment, however, the assessment instrument should be positioned explicitly as a measurement tool that requires systematic development, empirical testing, and sufficient validity evidence. Huang (2026) exemplifies this approach by developing and validating a process-based assessment scale prior to its use in evaluating learners' communicative performance. This approach highlights that simulation can serve a meaningful assessment function only when the evidence of learners' competence generated through the simulation is supported by a defensible and empirically grounded validity argument. In simulation-based ESP assessment, assessment instruments must be treated as measurement tools that need to be designed, tested, and supported by adequate validity evidence.

Context, Technology, and Gaps in the Evidence

Geographically, research on simulation-based ESP assessment is still largely developing within the contexts of East Asia and Southeast Asia. This situation holds strong relevance for vocational education in Indonesia because much of the research originates from contexts with similar characteristics, particularly in efforts to link learning to industry needs. However, this contextual proximity does not always indicate the maturity of the evidence, especially regarding the development and validation of assessment instruments. In

performance-based assessment, contextual relevance must go hand in hand with the quality of measurement evidence so that interpretations of learners' competencies can be justified (Kane, 2013). Thus, limitations in validation evidence are not only a methodological weakness but also the most critical area for development in ESP assessment research in Indonesia.

Recent developments indicate a growing interest in the use of simulation technologies—including virtual and artificial intelligence-based environments—in vocational education. These technologies open up opportunities to create work scenarios that are more complex, interactive, and closer to professional practice than text-based learning or conventional assignments (Chernikova et al., 2020; Nyström & Ahn, 2020). However, increased technological complexity does not automatically improve the quality of assessment. Instruments used in simulation environments still require evidence of validity, as the quality of the technology cannot replace the accuracy of the constructs being measured (Kane, 2013). Without an adequate validation process, innovations in AI or virtual simulation risk merely creating a new form of an old problem: the use of engaging activities that are unable to provide valid evidence of competency.

Implications

These findings have important implications for vocational assessment practices, policies, and research. At the practical level, the development of ESP assessments needs to shift the focus from merely using simulations as learning activities to using simulations as a source of evidence of performance. This means that learners are not only placed in simulated situations but are also assessed based on their ability to complete communication tasks that represent real-world job requirements. This principle aligns with the authentic assessment approach, which emphasizes the alignment between assessment tasks and the professional practices that constitute the learning objectives (Gulikers et al., 2004).

At the institutional and vocational education policy levels, greater attention is needed regarding standards for the development of language competency assessment instruments. An increase in scores following a learning intervention is not sufficient to demonstrate the success of an assessment if it is not accompanied by evidence regarding reliability, construct validity, and the appropriateness of score interpretation (Kane, 2013). Therefore, future research agendas should prioritize simulation-based assessment instruments as a primary focus, from the development stage and validity testing through to the transparent reporting of measurement evidence. In this regard, research such as that by Huang (2026) can be viewed as an early example of how simulations can be developed not only as learning environments but also as instruments for generating evidence of professional communication competencies.

LIMITATIONS

Several limitations should be considered when interpreting the findings. The review searched four databases, Lens.org, ERIC, DOAJ, and Garuda, but did not include Scopus, Web of Science, EBSCO, or ProQuest, which may have resulted in relevant studies being missed. The ERIC search covered 2017–2026 and may therefore have excluded studies published in 2016, while DOAJ yielded only one relevant record, possibly reflecting limited open-access literature or search limitations. In Garuda, the lack of a citation export function required some metadata to be processed manually. In addition, some records could not be accessed because of availability or access restrictions. These issues were treated as limitations of search coverage rather than grounds for exclusion based on the eligibility

criteria. Therefore, the findings should be interpreted within the scope of the databases, search period, and accessible records included in the review.

Second, the methodological quality was not considered when determining which studies to include, which is in line with the approach used in a scoping review. Therefore, the review includes studies that may have issues with their reliability, such as internal inconsistencies found in Abdullah and Soliman (2019) and Tyas et al. (2020). The results regarding effectiveness were not given more importance based on the quality of the studies. The final corpus consisted of eleven studies. This small number is also a significant result because it shows that truly simulation-based evaluation is still a restricted field of study. Finally, consistency in the charting process was ensured by using independent counts and verifying them during any changes made to the corpus. However, formal measures of inter-rater agreement, such as Cohen's kappa, were not computed. Therefore, some decisions in qualitative coding might still rely on the researcher's judgment.

CONCLUSION

This review maps the use of work-simulation-based ESP assessments in vocational education through eleven studies and identifies several key patterns. In previous research, work simulations were more commonly used as learning strategies or practice contexts—such as role-plays, authentic tasks, and virtual environments—while competency measurement was still frequently conducted through tests separate from the simulation activities. In terms of instrument quality, only a small fraction of the studies provided validation evidence for language proficiency measurement tools, while the others did not demonstrate a sufficient basis for validity. These findings indicate a significant gap in the development of simulation-based ESP assessments, namely the limited availability of instruments capable of authentically measuring professional communication performance while also being supported by robust measurement evidence. Therefore, future research should prioritize the development and validation of assessment instruments integrated with work simulations so that the evaluation of language competencies in vocational education does not stop at learning experiences that resemble the workplace but also produces accountable evidence of competency.

REFERENCES

- Abdullah, M. I. M., & Soliman, K. S. M. (2019). The Effectiveness of Using Task-Based Instruction in Enhancing Vocational Students' Communicative Performance. *Open Journal of Modern Linguistics*, 9(6), 446. <https://doi.org/10.4236/ojml.2019.96035>
- Anderson, H. L., Kurtz, J., & West, D. C. (2021). Implementation and Use of Workplace-Based Assessment in Clinical Learning Environments: A Scoping Review. *Academic Medicine*, 96(Supplement_2), S164–S174. <https://doi.org/10.1097/ACM.0000000000004366>
- Apriyanti, D., Syarif, H., & Ramadhan, S. (2021). Video Feature Making in ESP- Based Public Speaking Class: A Student-Centred Learning in Vocational Higher Education Context. *International Journal of Language Education*, 5(1). <https://doi.org/10.26858/ijole.v5i1.15419>

- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Centre for Reviews and Dissemination. (2025). *PROSPERO*. <https://www.crd.york.ac.uk/prospERO/>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Coxhead, A., Parkinson, J., Mackay, J., & McLaughlin, E. (2019). *English for Vocational Purposes: Language Use in Trades Education*. Routledge. <https://doi.org/10.4324/9780429449932>
- 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>
- Dahlback, J., Olstad, H. B., Sylte, A. L., & Wolden, A.-C. (2020). The Importance of Authentic Workplace-Based Assessment: A Study From VET Teacher Education. *International Journal for Research in Vocational Education and Training*, 7(3), 302–324. <https://doi.org/10.13152/IJRVET.7.3.3>
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–86. <https://doi.org/10.1007/BF02504676>
- Guo, J., & Yu, K. (2022). Effects of ESP Courses on Chinese Health Vocational College Students. *Korean J General Edu*, 16(5), 189–202. <https://doi.org/10.46392/kjge.2022.16.5.189>
- Hidayati, L., & Meisani, D. R. (2023). The Needs Analysis of English for Specific Purposes: A Study in an Indonesian Medical School. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 7(2), 465–479. <https://doi.org/10.21093/ijeltal.v7i2.1424>
- Huang, X. (2026). Reconstruction and Empirical Research on Full-Process Maritime English Teaching in Higher Vocational Colleges Supported by Dual Collaborative Generative AI Models. *International Journal of Education and Humanities*, 24(1), 55–59. <https://doi.org/10.54097/d28kdm66>
- Jossberger, H., Brand-Gruwel, S., van de Wiel, M. W. J., & Boshuizen, H. (2017). Learning in Workplace Simulations in Vocational Education: A Student Perspective. *Vocations and Learning*, 11(2), 179–204. <https://doi.org/10.1007/s12186-017-9186-7>
- Kanca, I. N., Ginaya, G., & Astuti, N. N. S. (2021). The effectiveness of the problem solving method on learning outcomes of the English course for room division operation during the COVID-19 pandemic. *International Journal of Linguistics, Literature and Culture*, 7(1), 12–22. <https://doi.org/10.21744/ijllc.v7n1.1102>
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>

- Knoch, U., & Macqueen, S. (2016). 18. Language assessment for the workplace. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment*. <https://www.degruyterbrill.com/document/doi/10.1515/9781614513827-020/html?lang=en>
- Koester, A. (2021). English for Occupational Purposes. In *The Encyclopedia of Applied Linguistics* (pp. 1–6). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405198431.wbeal0381.pub2>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 69. <https://doi.org/10.1186/1748-5908-5-69>
- Liao, L., Ye, S. X., & Yang, J. (2023). A mini review of communicative language testing. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1058411>
- Lo, Y.-H. G. (2017). Reconceptualizing Curriculum Politics: A Case Study of an ESP Program for Vocational High School Students in Taiwan. *Taiwan Journal of TESOL*, 14(1), 37–86.
- Messick, S. (1989). Validity. In *Educational Measurements* (3rd ed.). American Council on Education.
- Mukadimah, H., & Sahayu, W. (2021). Embedding Content and Language Integrated Learning (CLIL) into English for specific purposes curriculum for vocational high school. *Jurnal Pendidikan Vokasi*, 11(2), 204–212. <https://doi.org/10.21831/jpv.v11i2.41065>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational & psychological assessment*. SAGE Publications.
- Norris, J. M., Sasayama, S., & Kim, M. (2023). Simulating Real-World Context in an Email Writing Task: Implications for Task-Based Language Assessment. *ETS Research Report Series*, 2023(1), 1–22. <https://doi.org/10.1002/ets2.12366>
- Nurjanah, I., & Ana, A. (2022). Work Readiness of TVET Graduates in the Context of Industry 4.0. *Proceedings of the 4th International Conference on Innovation in Engineering and Vocational Education (ICIEVE 2021)*, 34–38. <https://www.atlantispress.com/proceedings/icieve-21/125972285>
- Nyström, S., & Ahn, S. (2020). Simulation-based training in VET through the lens of a sociomaterial perspective. *Nordic Journal of Vocational Education and Training*, 10(1), 1–17. <https://doi.org/10.3384/njvet.2242-458X.201011>
- OECD. (2021). Teachers and Leaders in Vocational Education and Training. *OECD*. https://www.oecd.org/en/publications/teachers-and-leaders-in-vocational-education-and-training_59d4fbb1-en.html
- Peters, M., Marnie, C., Tricco, A., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C., & Khalil, H. (2021). Updated methodological guidance for the conduct

- of scoping reviews. *JBI Evidence Implementation*, 19(1), 1–10. <https://doi.org/10.1097/XEB.0000000000000277>
- Plomp, T., & Nieveen, N. (2013). *Educational Design Research: Part A: An Introduction*. https://ris.utwente.nl/ws/portalfiles/portal/1554416/Design_Research_Part_A.pdf
- Pradipta, B., Hirawan, F., & Ragamustari, S. (2021). Evaluation of policy in the vocational education system revitalization in Indonesia: Examining the teaching factory readiness of industry. *Jurnal Pendidikan Vokasi*, 11(1), 68–77. <https://doi.org/10.21831/jpv.v11i1.37693>
- Prianto, A., Winardi, W., & Qomariyah, U. N. (2021). The Effect of the Implementation of Teaching Factory and Its Learning Involvement toward Work Readiness of Vocational School Graduates. *International Journal of Instruction*, 14(1), 283–302. <https://doi.org/10.29333/iji.2021.14117a>
- Putra, I. M. A., Widanta, I. M. R. J., Ardika, I. W. D., Hudiananingsih, P. D., Yuliantini, N. N., & Sitawati, A. A. R. (2021, April 26). Designing English for Vocational Purposes (EVP) in the Context of Merdeka Belajar of Kampus Merdeka. *Proceedings of the International Conference on Applied Science and Technology on Social Science (ICAST-SS 2020)*. International Conference on Applied Science and Technology on Social Science. <https://doi.org/10.2991/assehr.k.210424.051>
- Rafiq, K. R. M., Hashim, H., & Yunus, M. M. (2021). Sustaining Education with Mobile Learning for English for Specific Purposes (ESP): A Systematic Review (2012–2021). *Sustainability*, 13(17), 1–18. <https://doi.org/10.3390/su13179768>
- Rahmanu, I. W. E. D., & Molnár, G. (2026). Virtual reality in English language education: A systematic review of opportunities and challenges in higher education. *Acta Psychologica*, 264, 106358. <https://doi.org/10.1016/j.actpsy.2026.106358>
- Ramendra, D. P., Juniarta, P. A. K., Parma, I. P. G., Jayanta, I. N. L., Tantri, A. A. S., & Dewantara, K. A. K. (2025). Artificial Intelligence-Based Virtual Tour for Vocational High Schools in Tourism Sector in Developing English Language Competence for Guides. *International Journal of Language Education*, 9(1). <https://doi.org/10.26858/ijole.v1i1.71704>
- Rattanavich, S. (2017). The Effects of an Experimental Training Program for Teachers of Vocational English Using Concentrated Language Encounter Instructional Processes and Reciprocal Peer Teaching Strategies. *English Language Teaching*, 10(12), p58. <https://doi.org/10.5539/elt.v10n12p58>
- Rietmeijer, C. B. T., Watling, C. J., & Teunissen, P. W. (2025). Rethinking Workplace-Based Assessment: The Costly Illusion of Authenticity. *Academic Medicine*, 100(11), 1254–1257. <https://doi.org/10.1097/ACM.00000000000006173>
- Saputro, D. A., & Savitri, W. E. (2025). Need Analysis on English for Specific Purposes of Visual Communication Design Student at Vocational High School. *Journal of Linguistics, Literature, and Language Teaching (JLLLT)*, 4(2), 139–151. <https://doi.org/10.37249/jlllt.v4i2.892>
- Sari, L. I., Sari, R. H., & Amrullah, Y. F. (2024). Empowering ESP Learners Through Business Meeting Simulations: Insights on Challenges and Opportunities. *English Review: Journal of English Education*, 12(2), 537–546. <https://doi.org/10.25134/erjee.v12i2.9267>

- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education*, 7(1), 1719571. <https://doi.org/10.1080/2331186X.2020.1719571>
- Su, S.-M., Tsai, Y.-H., & Tai, H.-C. (2022). An ESP Approach to Teaching Nursing Note Writing to University Nursing Students. *Education Sciences*, 12(3), 223. <https://doi.org/10.3390/educsci12030223>
- Suharno, Pambudi, N. A., & Harjanto, B. (2020). Vocational education in Indonesia: History, development, opportunities, and challenges. *Children and Youth Services Review*, 115, 105092. <https://doi.org/10.1016/j.childyouth.2020.105092>
- Sumira, Febriansyah, G., Miladiyenti, F., Wahyuni, N., & Laily, H. I. (2025). Simulation-Based Project Learning in ESP: The Effectiveness of Crewpers for Vocational Journalism Education. *Voices of English Language Education Society*, 9(2), 400–412. <https://doi.org/10.29408/veles.v9i2.30694>
- Tricco, A. C., Lilie, E., Zarin, W., O'Brien, K., Colquhoun, H., Levac, D., Moher, D., Peters, M., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M., Garritty, C., ... Straus, S. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Tyas, A. S. P., Muam, A., Sari, Y. I. H., & Dewantara, C. (2020). The Effectiveness of Blended Learning in Improving Students' Workplace Communication Skills: A Case Study on Olive Website Test Result. *Lingua Cultura*, 14(1), 1–13. <https://doi.org/10.21512/lc.v14i1.6130>
- Xenia, T., & Indriyani, C. E. (2025). Hoteliers' Needs Analysis in a Five-Star Hotel: ESP Course Design Considerations. *Journal of English Language and Education*, 10(6), 212–222. <https://doi.org/10.31004/jele.v10i6.1563>
- Zinchenko, O., Ogrenich, M., Shepel, M., & Yakovlieva, M. (2020). Future Economists' Cultural Competence and Business English Speech Etiquette Formation. *Romanian Journal for Multidimensional Education/Revista Românească Pentru Educație Multidimensională*, 12(2), 290. <https://doi.org/10.18662/rrem/12.2/279>

APPENDIX 1. SEARCHING STRATEGY

All searches were conducted on July 4, 2026. The publication period was set to 2016–2026 for Lens.org, DOAJ, and Garuda; ERIC was restricted to 2017–2026 because its interface only provides pre-set year filters. The search strings are presented verbatim as entered into each interface.

Table L1. Summary of Searches by Database

Database	Field and Filters	Year Range	Result
Lens.org	All fields; journal articles; DOI/CORE/OpenAlex sources	2016–2026	115
ERIC	Keyword; “Full text available on ERIC” selected	2017–2026	44 ¹
DOAJ	All field	2016–2026	1
Garuda	Title field	2016–2026	12 ²
Total Identified			172

¹ ERIC used three separate search strings (9 + 11 + 24 = 44 raw records); overlaps among the strings were removed during the deduplication stage.

² the number of Garuda articles successfully downloaded (Garuda does not support BibTeX export; some results could not be downloaded due to webpage access issues), not the number of articles that met the eligibility criteria.

Following the compilation of 172 records, 12 duplicates were identified and removed, resulting in 160 unique records retained for title and abstract screening.

Lens.org

Date: 4 July 2026 · **Filter:** 2016–2026, journal article, source; DOI/CORE/Open Alex
Result: 115

("English for Specific Purposes" OR (ESP OR ("English for Occupational Purposes" OR ("vocational English" OR "workplace English")))) AND (("vocational education" OR (TVET OR ("vocational students" OR ("technical education" OR "technical and vocational education and training")))) AND (("simulation-based learning" OR ("workplace simulation" OR ("role play" OR ("authentic task" OR ("task-based learning" OR ("experiential learning" OR ("workplace scenario" OR "situated learning")))))) AND (assessment OR ("performance assessment" OR ("authentic assessment" OR ("competency assessment" OR ("professional communication" OR ("workplace communication" OR "communication competence"))))))))

ERIC (eric.ed.gov)

Date: 4 July 2026 · **Filter:** *Full text available on ERIC* Selected; Peer reviewed only was left unchecked · **Year Range:** *Since 2017* (2017–2026) · **Results (raw):** 44 = 9 + 11 + 24

String 1 — 9 results

("Vocational Education" OR TVET OR "Technical and Vocational Education and Training" OR "vocational students") AND ("English for Specific Purposes" OR ESP OR "English language learning" OR "English language") AND ("simulation" OR "role play" OR "mock interview" OR "work-based learning" OR "experiential learning" OR "authentic assessment" OR "authentic task")

String 2 — 11 results

("Vocational Education" OR TVET OR "vocational students") AND ("English" OR "English language" OR "English for Specific Purposes" OR ESP) AND ("simulation"

OR "role play" OR "work-based learning" OR "experiential learning" OR "authentic task")

String 3 — 24 results

("English" OR "English for Specific Purposes" OR ESP) AND ("Vocational Education" OR TVET OR "vocational students") AND (assessment OR "performance assessment" OR "authentic assessment" OR "competency assessment")

ERIC Note. The search was divided into three separate strings because the ERIC interface limits the complexity of individual queries. As only preset year filters were available, the ERIC search was restricted to 2017–2026, potentially excluding ERIC-indexed publications published in 2016.

DOAJ

Date: 4 July 2026 · **Result:** 1

("English for Specific Purposes" OR "vocational English") AND (simulation OR "role play" OR "task-based") AND (assessment OR "communication skills")

Garuda (garuda.kemdiktisaintek.go.id)

Date: 4 July 2026 · **Field:** Title · **Year Range:** 2016–2026 · **Downloaded:** 12

No.	Field Title	Hasil
1	"English for Specific Purposes" AND "pendidikan vokasi"	1
2	"ESP simulasi vokasi" AND "bahasa Inggris kejuruan asesmen"	0
3	"ESP simulasi vokasi"	5
4	"bahasa Inggris kejuruan asesmen"	0
5	"English vocational assessment"	16
6	"English learning" AND "vocational" AND "simulation"	1

Garuda Note. The search was restricted to the *Title* field; relevant records in which the search terms appeared only in abstracts or keywords may therefore have been missed. As Garuda does not support citation export in BibTeX format, metadata were manually retrieved for each article. Of the total search results, 12 articles were successfully downloaded and included in the deduplication stage. The remaining records could not be retrieved due to website errors or access restrictions encountered on July 4, 2026. Records that could not be downloaded were not assessed for eligibility because their full texts were inaccessible. This limitation reflects a retrieval constraint rather than exclusion based on the eligibility criteria and is acknowledged as a potential source of coverage bias affecting the representation of national publications.