Journal of English Language and Education
ISSN 2597- 6850 (Online), 2502-4132 (Print)
Journal Homepage: https://jele.or.id/index.php/jele/index

OPEN ACCESS

Article

Course Assessment Patterns during Collaborative ESP Reading among Management Students at an Indonesian University

https://doi.org/10.31004/jele.v11i6.2601

Erwin Suhendra, Lalu Arwan Darmajaya[ab]
[1]Universitas Bumigora, Indonesia
[2]Politeknik Kesehatan Mataram, Indonesia
Corresponding Author: erwin@universitasbumigora.ac.id

A B S T R A C T

This study descriptively examined course-based ESP assessment patterns during collaborative reading implementation among 171 Management students in four intact Bahasa Inggris 2 classes at an Indonesian university. The activities used jigsaw reading, peer explanation, vocabulary mapping, collaborative summarising, and group discussion in groups of four to six students. Using a classroom-based descriptive-comparative design, the analysis compared institutional MID-TERM TEST and FINAL TEST records and examined class-level change, pass rates, grade distributions, and lower-grade component profiles. The assessments were sequential course indicators rather than equivalent parallel forms: both included reading comprehension, vocabulary and grammar, and short writing, while the final assessment also included a group presentation and report. The overall mean increased descriptively from 68.71 to 75.18, an absolute change of 6.47 points (9.42% relative to the mid-term mean). Class-level changes ranged from 3.51 to 9.03 points. The recorded pass rate was 93.57%, and 148 of 171 students received A or B+; however, lower-grade profiles included seven D and four E grades, commonly co-occurring in the records with zero, very low, or imbalanced assessment components. No inferential test or standardized effect size was reported, and the one-group institutional-record design does not establish that collaborative reading caused the observed differences. The study shows how course records can support diagnostic monitoring of ESP assessment patterns across intact classes while also revealing the limits of such records for causal explanation.

Keywords: *Collaborative ESP Reading, Course-Based Assessment, Institutional Records, Diagnostic Monitoring, Management Students, Indonesian Higher Education*

Article History:
Received 06th October 2026
Accepted 31st October 2026
Published 07th November 2026

Check for updates

INTRODUCTION

English for Specific Purposes (ESP) literacy is increasingly relevant for management students in Indonesian higher education because disciplinary learning now requires regular engagement with English texts, business terminology, digital information, and professional genres. Management students are expected to read case materials, interpret market-related information, understand workplace documents, and access academic resources that frequently circulate in English. In this context, ESP literacy functions as disciplinary literacy because students must understand how meanings are organized in texts connected to their field, profession, and future workplace participation. ESP instruction therefore needs to connect general language learning with the communicative demands of a specific academic and professional domain (Hyland, 2019). Stoller & Robinson (2018) argue that ESP teaching should be grounded in discipline-relevant practices and materials, while Hirvela (2012, 2026) positions ESP reading as one important route through which learners enter specialized discourse communities. For Indonesian management students, this issue is significant because English is closely linked with academic access, employability, and professional communication.

The construct examined in this study is better understood as integrated ESP literacy rather than reading comprehension alone. ESP reading usually involves identifying main

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

ideas, making inferences, understanding vocabulary in context, and interpreting genre-based meanings (Dai, 2024; Duong & Le Truong, 2022; Hirvela, 2025). However, assessment in a university English course may also include grammar, vocabulary development, summary writing, project work, presentation, and professional communication tasks. Integrated ESP literacy therefore refers to students' ability to use English across connected academic and professional tasks, including reading texts, understanding technical vocabulary, applying language forms, producing written responses, and communicating field-related meanings (Kinnear, 2018; Synekop & Lytovchenko, 2024). This construct is appropriate for the present study because the institutional records include daily score, mid-term test, final test, final score, and grade, while the final institutional weighting formula was treated according to the official recorded scores rather than reconstructed mathematically.

A persistent challenge in ESP classrooms is the heterogeneity of student proficiency. Non-English-major students often enter university English courses with different levels of prior exposure, vocabulary knowledge, grammar control, reading confidence, and familiarity with academic or professional genres. Shen (2013) reports that technical university learners experience academic reading difficulties because disciplinary texts require language knowledge, strategy use, and content familiarity. Chinpakdee (2024) also shows that university students' reading difficulties are connected to lexical demands, reading-related skills, and teacher expectations. These challenges become more visible in management classrooms where students must process business-related texts with varied preparation. Passive classroom routines may further reduce opportunities for students to clarify meaning, test interpretations, and build vocabulary through interaction. A lecturer-fronted reading lesson may provide structure, but it may also limit peer explanation, shared comprehension, and active negotiation of meaning.

Collaborative learning offers a structured pedagogical response to these classroom conditions. In reading instruction, collaboration can be organized through jigsaw reading, group discussion, peer explanation, vocabulary mapping, collaborative summarising, role distribution, lecturer facilitation, and feedback (Hatte et al., 2025; Shaik, 2025; Suparto et al., 2025). These procedures make reading more interactive because students work with peers to divide text responsibilities, explain content, clarify difficult terms, ask comprehension questions, and reconstruct meaning. Collaborative Strategic Reading and jigsaw reading have been widely discussed as approaches that combine reading strategy instruction with cooperative learning procedures (Aldossary, 2024; Butros Abed, 2019; Khampool & Chumworatayee, 2023). In ESP and EAP contexts, collaborative reading is especially relevant because learners must process both language and disciplinary content. The lecturer remains central because structured collaboration requires task design, grouping, monitoring, role allocation, and debriefing.

Previous studies support the pedagogical value of collaborative reading, although their evidence should be interpreted in relation to each study design. Butros Abed (2019) examined jigsaw as a cooperative learning strategy for EFL reading comprehension and reported post-intervention differences between experimental and control groups. Khampool & Chumworatayee (2023) found that Collaborative Strategic Reading supported Thai EAP university learners' reading comprehension across proficiency levels, particularly in main idea, vocabulary in context, inference, and referential relationships. Aldossary (2024) reported that Collaborative Strategic Reading was associated with improved reading comprehension and positive learner responses among EFL university students. More recent studies have also expanded collaborative reading into technology-supported and diagnostic contexts. Fan & Yan (2025) examined diagnosis-based pedagogical intervention and found that peer-collaborative and teacher-guided groups developed several micro reading abilities across repeated diagnostic assessments. Gaffas (2024) showed that microblogging-based ESP instruction supported specialized-language performance, reading and writing perceptions, and collaborative discussion in a medical ESP course. These studies indicate growing interest in collaborative reading, peer explanation, vocabulary development, and ESP literacy.

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

Peer explanation and vocabulary negotiation are central mechanisms in collaborative ESP literacy. When students explain a paragraph, define a technical term, or justify an interpretation to classmates, they reorganize understanding into communicable form. This process can make comprehension visible and open to correction. In ESP classrooms, peer explanation also helps students connect language with disciplinary meaning because technical vocabulary often carries specialized meanings that cannot be understood through general English knowledge alone (Hassan et al., 2023; Suparto et al., 2025; Vasconcelos, 2023). Studies on collaborative academic reading indicate that students co-construct comprehension through paraphrasing, elaboration, questioning, and summarising (Jin et al., 2020; Liu et al., 2024). Collaborative vocabulary mapping can also support comprehension by requiring students to identify difficult terms, discuss possible meanings, relate vocabulary to context, and build shared understanding. In this study, vocabulary was also part of the assessed construct because both the mid-term test and final test included vocabulary and grammar sections supported by answer keys and scoring rubrics. Therefore, vocabulary development is treated as both a pedagogical pathway within collaborative reading and a documented component of course-based integrated ESP literacy assessment.

The effectiveness of collaborative learning depends on structure. Placing students in groups does not automatically create productive interaction, equal participation, or reliable learning evidence. Group formation, role distribution, individual accountability, and lecturer facilitation shape whether collaborative tasks become purposeful or remain superficial. Research on collaborative learning emphasizes that group dynamics, role clarity, and accountability influence the quality of interaction and group outcomes (Han et al., 2025; Volet et al., 2017). In language classrooms, roles such as summariser, questioner, vocabulary checker, and timekeeper are pedagogically useful because they connect classroom interaction with specific literacy processes: identifying main ideas, asking comprehension questions, clarifying technical terms, and managing task completion. Lecturer facilitation is equally important because collaborative learning requires clear instructions, task modelling, monitoring, feedback, and debriefing. Chang-Tik (2023) links lecturers' approaches with the quality of group-based learning, while studies of online and face-to-face collaboration highlight the need for procedural support, feedback, and attention to unequal participation (Bin Dahmash, 2025; Butarbutar et al., 2023).

Assessment is a central issue in this study because the available evidence comes from institutional course records rather than a standardized reading test. Authentic classroom assessment can provide ecological evidence because it reflects how students perform in a real course with actual scoring components. In ESP and EAP contexts, such evidence is useful because learning is often distributed across reading, vocabulary work, writing, participation, projects, and examinations. However, authentic assessment also requires careful interpretation. Course records may show what was recorded and graded, but they do not automatically identify the exact construct measured unless the assessment blueprint, task format, scoring rubric, and weighting formula are available. A reading-only study can report ESP reading achievement if the test clearly measures reading comprehension skills. A broader course assessment is more safely described as integrated ESP literacy performance or course-based ESP assessment performance, especially when daily score, mid-term test, final test, final score, and grade are involved.

Despite growing research on collaborative reading, several gaps remain. First, many studies emphasize aggregate improvement in reading comprehension, while fewer studies examine how achievement patterns vary across intact ESP classes. Second, classroom intervention studies often use pre-test and post-test scores without sustained attention to grade distribution, pass rates, and lower-score profiles. Third, ESP collaborative reading research gives limited attention to institutional assessment records as authentic classroom evidence. Such records can reveal patterns that a single mean score may hide, including class-level variation and weak or missing assessment components. This assessment gap matters because real ESP courses are commonly evaluated through daily work, mid-semester examinations, final examinations, final scores, and grades. Stronger collaborative reading

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

studies usually clarify treatment duration, test format, skill coverage, and analysis procedures (Aldossary, 2024; Fan & Yan, 2025; Khampool & Chumworatayee, 2023). The present study uses a different evidentiary logic by focusing on diagnostic grade-profile evidence.

Taken together, the literature narrows the problem from the broad need for disciplinary ESP literacy to a specific evidentiary question: how can authentic course records be used to describe assessment patterns across intact management classes during structured collaborative reading? Answering this question requires more than an overall mean because aggregate change can obscure between-class variation, the distribution of final grades, and component-level vulnerability among lower-performing records. It also requires explicit recognition that the mid-term and final assessments are sequential, non-equivalent course indicators.

This study examines course-based ESP assessment performance during the implementation of collaborative ESP reading activities in four intact Bahasa Inggris 2 classes in the Management Study Program at Universitas Bumigora, Indonesia. It uses a classroom-based descriptive-comparative assessment design based on a one-group mid-term test–final test score comparison and diagnostic analysis of institutional course records. The dataset includes 171 students across four classes: Class A (n = 40), Class B (n = 45), Class C (n = 45), and Class D (n = 41). The analysis focuses on mid-term test–final test score change, class-level gain variation, pass rates, grade distribution, and lower-grade component patterns. The study does not position itself as strong experimental proof of collaborative learning. Instead, its originality lies in using authentic institutional records diagnostically to examine how collaborative ESP reading implementation is reflected in course assessment outcomes across parallel intact classes. Its contribution is methodological, pedagogical, and institutional because it shows how classroom records can support cautious monitoring of integrated ESP literacy performance while highlighting the need for assessment transparency, intervention documentation, and component-level support.

Accordingly, the study asks what descriptive change is visible between the recorded mid-term and final-test indicators, how the recorded gains and pass rates vary across the four intact classes, and what the final grade distribution and lower-grade component profiles reveal for diagnostic monitoring. These questions concern recorded assessment patterns; they do not test whether collaborative reading caused those patterns.

The study therefore aims to provide a transparent, non-causal diagnostic account of those assessment patterns. To address the questions, the analysis moves from cohort-level mid-term–final score change to class-level gain and pass-rate variation, then to final-grade distribution and lower-grade component profiles. The remainder of the manuscript reports the classroom-based design, setting, participants, implementation, assessment, and analysis procedures before presenting and interpreting these layers of evidence and drawing bounded pedagogical implications.

METHOD

This study used a classroom-based descriptive-comparative assessment design with a one-group mid-term test-final test structure. In this study, descriptive-comparative means that recorded means, distributions, pass rates, and component patterns were compared without random assignment, a control group, or inferential causal testing. The design was used because the study was conducted in an existing Bahasa Inggris 2 course with four intact classes. Random assignment, a control group, and full experimental control were unavailable in this setting. The mid-term test served as the mid-semester assessment indicator, while the final test served as the final-semester assessment indicator. Thus, the findings are interpreted as institutional assessment patterns during the collaborative ESP reading implementation period, not as causal evidence of instructional effectiveness.

The unit of analysis was the individual student's institutional assessment record, considered descriptively within one of four intact classes. The study was retrospective in its use of data because the completed course records were analysed after the assessment process,

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

while its temporal comparison comprised the recorded mid-semester and final-semester indicators for the same cohort. The first research question was linked to cohort-level mean change, the second to class-level gains and pass rates, and the third to grade frequencies and lower-grade component profiles. This RQ–data–analysis alignment limits the permitted inference to description, comparison, and diagnosis within the recorded cohort.

Research Site and Participants

The study was conducted in the Management Study Program at Universitas Bumigora, Indonesia. The course was Bahasa Inggris 2, offered in the even semester of the 2022/2023 academic year. The course formed part of the English provision for management students and was designed to support students' ability to use English for academic and professional communication related to business, management, tourism, small enterprise activity, and workplace interaction.

The participants were 171 students from four intact classes. Class A consisted of 40 students, Class B consisted of 45 students, Class C consisted of 45 students, and Class D consisted of 41 students. These classes were retained as natural classroom groups because the study used existing course records. The dataset consisted of institutional assessment records and did not include systematic demographic information, previous English scores, attendance records, motivation scores, or classroom observation data. Therefore, the analysis focused only on assessment components and class-level patterns visible in the institutional records.

Those four intact classess represented all Bahasa Inggris 2 subject taught in the 2022/2023 even semester. The four intact classes constituted the accessible Bahasa Inggris 2 cohort for the semester, so inclusion operated as a census of the available class records rather than as probability or purposive sampling. The analysis therefore did not use participant recruitment, random selection, or an a priori power calculation. The sample size was determined by the institutional cohort represented in the available records, and the scope of inference is correspondingly bounded to this course and comparable contexts rather than to the wider population of Indonesian management students.

Collaborative ESP Reading Implementation

The classroom implementation consisted of structured collaborative ESP reading activities embedded in regular Bahasa Inggris 2 instruction. It used jigsaw reading, peer explanation, vocabulary mapping, collaborative summarising, and group discussion to help students process management-related texts through shared comprehension, vocabulary clarification, and group task completion. These activities were aligned with the course objective of developing students' English for academic and professional communication in business, tourism, small enterprise activity, and workplace contexts.

Each session followed a repeated sequence. The lecturer introduced the topic, linked the text to management-related situations, and divided students into heterogeneous groups of four to six members. Each group identified main ideas, clarified unfamiliar vocabulary, discussed factual information, inferred simple meanings, and produced a short oral or written summary. The lecturer then led a whole-class debriefing to confirm answers, clarify vocabulary, correct misunderstanding, and connect the activity with course outcomes.

Students were assigned functional roles to make participation more visible. The summariser restated key ideas, the questioner raised comprehension or inference questions, the vocabulary checker clarified difficult terms, and the timekeeper maintained task progress. These roles connected group interaction with reading comprehension, vocabulary use, sentence-level meaning, short written responses, and oral explanation.

The lecturer facilitated the classroom implementation by explaining procedures, modelling reading strategies, monitoring group work, giving feedback, and checking whether students could justify answers from the text. The activities were linked to the course assessment: peer explanation and jigsaw reading supported reading comprehension, vocabulary mapping supported vocabulary and grammar, collaborative summarising

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

supported the short essay task, and group discussion supported the final project presentation and report.

The collaborative ESP reading implementation was conducted over 12 weekly sessions in each of the four intact classes. Each class received one 90-minute session per week over a 12-week implementation period. Thus, students in each class received a total of 1.080 minutes, equivalent to 18 hours of collaborative ESP reading instruction. Across the four classes, the implementation involved 48 scheduled class meetings, representing 4.320 minutes or 72 hours of classroom delivery in total.

The same instructional sequence was scheduled for all four classes, comprising lecturer orientation, small-group text processing, role-based collaboration and whole-class debriefing. During each session, students worked with ESP reading materials related to business, management, tourism, small enterprise activity and workplace communication. The repeated weekly structure was intended to provide consistent opportunities for collaborative comprehension, vocabulary development, summarising, peer explanation and discussion throughout the implementation period. The available institutional dataset contained assessment records rather than observation logs, attendance records, session-by-session fidelity documentation or records of classroom deviations. Consequently, the analysis of implementation was based on the scheduled instructional protocol and the assessment patterns recorded for the four classes.

Across classes, the reported implementation schedule comprised 12 sessions delivered at one meeting per week, with the same sequence of lecturer orientation, small-group processing, role-based collaboration, and whole-class debriefing. This description specifies the intended common instructional protocol. However, the dataset did not include observation logs, attendance data, session-by-session fidelity records, or documented deviations, so equivalence of delivery and participation across classes could not be independently assessed.

Assessment Model and Construct Rationale

The study used course-based integrated ESP literacy assessments as the main instruments. The mid-term and final tests were aligned with the Bahasa Inggris 2 course, the Management Study Program context, and the English Unlimited A2 Elementary syllabus. The mid-term test covered Units 1–7, lasted 90 minutes, and consisted of reading comprehension, vocabulary and grammar, and a short essay task, with a total written score of 30 points. The final test covered Units 8–12, lasted 120 minutes, and included the same written components plus a group project presentation and report.

The assessment texts were contextualized for management students. The mid-term reading text focused on a campus small-business project involving product promotion, customer communication, pricing, and teamwork. The final reading text used a tourism-business context involving Bali, small businesses, tourism services, customer service, and English use in business communication. Therefore, the construct examined in this study is integrated ESP literacy rather than pure reading comprehension. The tests assessed reading comprehension, vocabulary and grammar, short written production, and, in the final test, collaborative professional communication through project presentation and report.

The mid-term and final tests were not equivalent parallel forms: they covered different units, and the final assessment added a group project presentation and report. The two scores are therefore treated as sequential course indicators rather than directly comparable measures from an unchanged instrument.

The variables used in the reported analyses were class membership, mid-term test score, final-test score, final score, and grade. The individual record was the measurement unit, class was the grouping variable, and the two test occasions were ordered institutional indicators. Because content coverage and task composition changed between occasions, the analysis did not assume measurement invariance or interpret the mean difference as change on an identical standardized scale.

Scoring Rubric and Assessment Components

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

The mid-term test consisted of 10 reading comprehension items, 10 vocabulary and grammar items, and one short essay task. Each objective item was scored 1 point for a correct answer and 0 points for an incorrect, irrelevant, or blank answer. The essay was scored using four criteria: content relevance, organization, vocabulary use, and grammar accuracy. Content relevance had a maximum score of 4 points, while the other three criteria had 2 points each. The total written score was 30 points.

The final test followed the same written-test structure and added a group project presentation and report. The project rubric assessed content and relevance, ESP vocabulary use, report organization, oral presentation clarity, collaboration and role distribution, and visual or media support, with a total score of 20 points. Institutional records included daily score, mid-term test score, final test score, final score, and grade. The grade scale was A = 81–100, B+ = 71–80, B = 66–70, C+ = 61–65, C = 50–60, D = 30–49, and E = 0–29. The study used the official recorded scores and did not reconstruct final scores mathematically.

Institutional records included daily score, mid-term test score, final test score, final score and grade. The institutional grading scale was A = 81–100, B+ = 71–80, B = 66–70, C+ = 61–65, C = 50–60, D = 30–49, and E = 0–29. In accordance with the university's academic policy, the minimum passing grade for the course was C, corresponding to a final score of 50. Accordingly, students with a final score of 50 or above (grades C, C+, B, B+, or A) were classified as having passed the course. This institutional threshold was used to calculate the pass rates reported in the study. The analysis used the official final scores and grades recorded in the institutional dataset.

Test Administration and Data Management

The mid-term test and final test were administered within the official course assessment schedule. The mid-term test functioned as the mid-semester assessment, while the final test functioned as the final-semester assessment. The written components were completed individually and assessed using the prepared answer key and scoring rubric. The final test also included a group project presentation and written report, which assessed collaborative and professional communication components of integrated ESP literacy.

The tests were aligned with the course RPS and were used as part of the regular Bahasa Inggris 2 assessment system. Permission to conduct the study was obtained from the Management Study Program. The study used institutional course records after the assessment process had been completed. Student names and identities were anonymized before analysis. Results were reported at aggregate, class, and anonymized profile levels. Lower-grade cases were interpreted only through recorded score patterns, such as mid-term test or final test scores recorded as zero or very low. The study did not infer personal reasons for these scores because attendance, motivation, health, anxiety, and engagement data were not included in the dataset.

Data preparation was constrained to what could be verified from the completed institutional records. Recorded zeros and very low scores were retained as observed values and were not assigned personal causes; no demographic, attendance, motivation, or observation variables were added. The available documentation does not establish whether duplicate-identifier checks, range checks, discrepancy reconciliation, or missing-value procedures were conducted before analysis. Accordingly, the study makes no claim about unreported cleaning or imputation steps, and the stated denominators should be read as the records included in each reported summary.

Data Analysis

The analysis was descriptive-comparative and diagnostic. First, the overall mean mid-term and final-test scores were calculated for 171 students, and the mean difference was obtained by subtracting the mid-term mean from the final-test mean. The relative change in the group means was calculated as 6.47 divided by 68.71 and multiplied by 100, yielding 9.42%. This percentage is a descriptive comparison of group means, not a standardized effect or a significance test. Second, the same mean-difference procedure was applied to each class to

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

identify class-level variation. Third, pass rates and grade frequencies were calculated from the institutional grade records.

Lower-grade profiles were then examined descriptively to identify visible component-level patterns among students who received C, D, or E grades. This diagnostic analysis focused on recorded patterns such as a zero mid-term test, a zero final test, a very low final-test score, or imbalanced component performance. No inferential statistics were reported. A standardized effect size was not calculated because the reviewed source documents do not report the paired-score dispersion statistics required for that calculation. The results are therefore interpreted as descriptive assessment patterns rather than statistically tested instructional effects.

The method–result correspondence is as follows: the overall mean comparison produces Table 1; class-specific means, gains, and pass rates produce Table 2; grade frequencies produce Table 3; and the diagnostic review of lower-grade components produces Table 4. For each output, interpretation follows the descriptive statistic or record pattern and remains separate from causal explanation. No significance threshold, confidence interval, multivariable model, or sensitivity analysis was applied because the study did not undertake inferential testing.

Validity, Reliability, Trustworthiness and Ethics

Content validity was supported through alignment among the *Bahasa Inggris 2* RPS, instructional content, assessment tasks, and management students' learning needs. The mid-term test covered Units 1–7 using a campus small-business context, while the final test covered Units 8–12 using a tourism-business context. Both assessments included reading comprehension, vocabulary and grammar, and short essay writing, with the final assessment also including a group project presentation and report.

Scoring consistency was supported by answer keys and analytic rubrics. Objective items were scored using correct-or-incorrect marking, while essay and project tasks were assessed using predefined criteria. Official institutional assessment records were used for the analysis, providing authentic evidence of students' performance in the regular course setting. Measurement quality was interpreted based on content alignment, scoring procedures, and the available course records.

Ethical approval for the secondary use of the assessment data was obtained from the institutional ethics committee. Permission to use the course records was also obtained from the Head of the Management Study Program. Students provided informed consent at the beginning of the course for the use of their course-related data for research purposes.

All student records were anonymized before analysis. The dataset was stored securely, access was restricted to the researcher, and the data were used solely for this study. Results were reported in aggregated or anonymized form to maintain student confidentiality.

FINDINGS AND DISCUSSION

Overall Mid-Term Test-Final Test Score Change

This combined journal section distinguishes the recorded numerical findings from their interpretation. Within each subsection, the values are presented first and are followed by a separately labelled interpretation. The analysis focuses on mid-term and final-test scores, final grades, and lower-grade score profiles. Because the dataset did not include attendance, motivation, engagement, observation, or interview data, interpretation is limited to the recorded assessment evidence. Inferential statistics were not used because the dataset was not designed as a controlled instructional study.

The first analysis compared the overall MID-TERM TEST and FINAL TEST scores for the full cohort of 171 students. MID-TERM TEST was treated as the mid-semester assessment indicator, while FINAL TEST was treated as the final-semester assessment indicator. Table 1 presents the overall score change.

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

Table 1. Overall MID-TERM TEST-FINAL TEST Score Change

N	Mean MID-TERM TEST	Mean FINAL TEST	Mean gain
171	68.71	75.18	+6.47

The group mean increased descriptively from 68.71 to 75.18, an absolute difference of 6.47 points and a relative difference of 9.42% from the mid-term mean. The overall increase from MID-TERM TEST to FINAL TEST suggests stronger course-based ESP assessment performance during the period in which collaborative ESP reading activities were implemented. This pattern is broadly consistent with previous studies showing that Collaborative Strategic Reading and jigsaw reading can support comprehension, vocabulary processing, and learner participation in EFL or EAP contexts (Aldossary, 2024; Butros Abed, 2019; Khampool & Chumworatayee, 2023). However, the present study differs from those studies because it used institutional course records rather than a controlled experimental test. The gain should therefore be interpreted as descriptive classroom evidence, not as causal proof of intervention effectiveness.

This finding is important for integrated ESP literacy because the assessment did not measure reading comprehension alone. The MID-TERM TEST and FINAL TEST included reading comprehension, vocabulary and grammar, and short essay components, while the FINAL TEST also included a group project presentation and report. The overall gain therefore suggests improvement in course-based ESP assessment performance during the collaborative reading period. The result supports the pedagogical argument that collaborative reading can create opportunities for students to clarify meaning, negotiate vocabulary, and organize ideas, but it should not be used to claim that all components improved equally. The institutional records show the direction of performance change, while the design limits the explanation of the precise mechanism behind the change.

Class-Level Gain and Pass Rate

The second analysis examined whether the overall MID-TERM TEST–FINAL TEST pattern appeared consistently across the four intact classes. Table 2 presents the class-level MID-TERM TEST mean, FINAL TEST mean, gain score, and pass rate. Combining gain and pass rate in one table makes the class-level pattern visible without presenting excessive raw score detail.

Table 2. Class-Level MID-TERM TEST-FINAL TEST Gain and Pass Rate

Class	N	Mean MID-TERM TEST	Mean FINAL TEST	Mean gain	Pass rate
Class A	40	69.47	78.50	+9.03	97.50%
Class B	45	65.36	73.22	+7.86	91.11%
Class C	45	71.18	76.67	+5.49	95.56%
Class D	41	68.93	72.44	+3.51	90.24%
Overall	**171**	**68.71**	**75.18**	**+6.47**	**93.57%**

Table 2 shows that all four classes recorded higher FINAL TEST means than MID-TERM TEST means, but the gain was not uniform. Class A recorded the largest gain, increasing from 69.47 to 78.50, while Class D recorded the smallest gain, increasing from 68.93 to 72.44. Class B and Class C also showed positive gains, but their gain sizes differed. The overall pass rate reached 93.57%, with all four classes exceeding 90%.

The class-level pattern shows positive descriptive assessment change in all intact classes during the collaborative ESP reading implementation period, but the magnitude of change varied. This variation matters because collaborative learning depends on more than the existence of group tasks. Previous research indicates that group dynamics, role clarity, individual contribution, and lecturer facilitation shape the quality of collaborative learning outcomes (Chang-Tik, 2023; Han et al., 2025; Volet et al., 2017). In this study, the same broad collaborative model was implemented across parallel classes, yet the recorded gains differed. The records cannot explain whether these differences were caused by participation quality, group cohesion, student preparation, or classroom interaction because such data were not

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

collected. They do, however, show that class-level monitoring is necessary when collaborative learning is implemented across multiple intact groups.

This interpretation has methodological and pedagogical implications. Methodologically, the class-level pattern warns against relying only on the overall mean because it can hide differences among classes. Pedagogically, it suggests that collaborative ESP reading needs consistent facilitation, task monitoring, and role reinforcement across parallel classes. Class A's higher gain and Class D's lower gain should not be interpreted as evidence of different student motivation without supporting data. Instead, they should be treated as diagnostic signals that lecturers need to examine implementation consistency, group task completion, and component-level assessment patterns during the course.

Grade Distribution

The third analysis examined the final grade distribution. This analysis was used to show how course outcomes were distributed across grade categories rather than relying only on mean scores. Table 3 presents the distribution of A, B+, B, C+, C, D, and E grades across the four classes.

Table 3. Final Grade Distribution by Class

Class	A	B+	B	C+	C	D	E	Total
Class A	23	12	0	2	2	0	1	40
Class B	21	20	0	0	0	1	3	45
Class C	27	13	0	1	2	2	0	45
Class D	17	15	3	1	1	4	0	41
Total	**88**	**60**	**3**	**4**	**5**	**7**	**4**	**171**

Table 3 shows that final grades were strongly concentrated in the A and B+ categories. Across the full cohort, 88 students received A and 60 students received B+, meaning that 148 of 171 students obtained one of the two highest grade categories. This pattern indicates broad course-level attainment. At the same time, lower-grade outcomes remained visible: 4 students received C+, 5 received C, 7 received D, and 4 received E. The distribution therefore gives a broader picture than the mean gain alone.

The diagnostic value of this distribution lies in its ability to show both attainment and vulnerability in the same dataset. If the analysis stopped at the overall gain, the study would mainly show a positive score trend. By adding grade distribution, the records show how that trend was distributed across students and classes. This matters for classroom-based ESP assessment because institutional records can help lecturers identify students who pass successfully, students who remain in lower-passing categories, and students whose final grades indicate serious component-level weakness. This interpretation aligns with diagnosis-based approaches to assessment, where repeated or component-level evidence is used to identify areas that require support rather than merely to produce final scores (Fan & Yan, 2025).

The A/B+ concentration also needs careful interpretation. It may indicate strong course-level performance, but it may also reflect the cumulative nature of institutional assessment, where daily scores, mid-term test scores, final test scores, project work, and final grades are combined. The manuscript therefore avoids treating high final grades as direct evidence that every literacy component improved. Instead, the distribution is used diagnostically to show the shape of course outcomes. This is a stronger use of institutional records because it acknowledges both the value and the interpretive limits of classroom assessment data. Accordingly, the reported pass rate may reflect both performance across cumulative course components and the institutional grade classification; it should not be treated as an independent measure of collaborative-reading effectiveness.

Lower-Grade Profiles and Component-Level Vulnerability

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

The fourth analysis examined lower-grade profiles diagnostically. Long individual student-level profile tables are not included in the main text because they would make the Results section resemble an administrative grade report. Instead, Table 4 summarizes the lower-grade patterns visible in the records. Full anonymized profiles may be placed in an appendix if required by the target journal or conference.

Table 4. Diagnostic Summary of Lower Grade Profiles

Profile category	Number of cases	Main recorded pattern	Interpretation boundary
E profiles	4	MID-TERM TEST and FINAL TEST were commonly recorded as 0; some cases had all components recorded as 0	The reason for zero scores cannot be inferred from the records.
D profiles	7	FINAL TEST was commonly very low or recorded as 0	The records show component-level vulnerability, not its cause.
C/C+ selected profiles	9	Mixed patterns included MID-TERM TEST recorded as 0, very low FINAL TEST, or moderate complete scores	These lower-passing profiles should be interpreted descriptively.

The lower-grade profiles show that broad attainment can coexist with component-level vulnerability. Although most students obtained A or B+, the records still included C/C+, D, and E profiles linked mainly to zero, very low, or imbalanced MID-TERM TEST and FINAL TEST components. This finding extends the analysis beyond average gain. It shows that institutional assessment records can identify students who may require targeted support before final grades are completed.

This diagnostic use of assessment records is consistent with the logic of diagnosis-based intervention, where repeated evidence is used to identify specific areas of learning difficulty (Fan & Yan, 2025). In the present study, the evidence is not a diagnostic micro-skill test but institutional score components. This distinction is important. The records can show where vulnerability appears, but they cannot explain why it appears. For example, a zero score may reflect absence, non-submission, administrative recording, low performance, or another factor, but the dataset does not contain evidence to identify the reason. For this reason, the study interprets lower-grade profiles only as component-level vulnerability.

The lower-grade profiles also qualify the interpretation of collaborative learning. Collaborative ESP reading may support many students through peer explanation, vocabulary mapping, and shared comprehension, but it does not remove the need for individual assessment completion and component-level monitoring. This is where individual accountability becomes important. Volet et al. (2017) argue that individual contributions shape the quality of collaborative learning outcomes, while research on group-based learning highlights the role of lecturer support and procedural clarity (Chang-Tik, 2023; Han et al., 2025). The lower-grade profiles in this study reinforce that point: collaborative learning should be accompanied by mechanisms that help lecturers identify students whose recorded components show risk before the course ends.

Pedagogical Implications for Collaborative ESP Teaching

The findings suggest several implications for ESP lecturers who use collaborative reading in intact university classes. First, collaborative ESP reading should be implemented as a structured pedagogy rather than as informal group work. Roles such as summariser, questioner, vocabulary checker, and timekeeper need to be explained, monitored, and connected to assessment tasks. This structure helps students understand how group interaction supports reading comprehension, vocabulary use, short writing, and oral communication.

Second, institutional assessment records should be used formatively as well as summatively. The present study shows that MID-TERM TEST and FINAL TEST scores, grade distribution, and lower-grade profiles can reveal patterns that are not visible in the overall

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

mean. Lecturers can use such records to identify classes that require more support, groups that need clearer task monitoring, and students whose assessment components show early signs of vulnerability. This is especially important in heterogeneous ESP classrooms, where students may differ in vocabulary knowledge, confidence, and familiarity with disciplinary texts.

Third, assessment transparency should accompany collaborative learning. Students need to understand how reading comprehension, vocabulary and grammar, essay writing, and project presentation are assessed. Transparent rubrics can help students connect collaborative classroom activities with course outcomes. For example, vocabulary mapping can be linked to vocabulary and grammar items, collaborative summarising can support short essay writing, and group discussion can prepare students for project presentation and report writing. This alignment makes collaborative ESP reading more purposeful and reduces the risk that group work becomes disconnected from assessment.

Finally, the findings suggest that remedial support should be component-based. Students with weak or missing MID-TERM TEST or FINAL TEST components require targeted assistance, but the type of support should be based on documented evidence. A student with weak vocabulary and grammar may need focused language practice, while a student with an incomplete project component may need task-completion support. The study cannot identify the personal causes behind low scores, but it shows how institutional records can help lecturers locate where support is needed.

Summary of Result Patterns

Taken together, the findings show that institutional assessment records can reveal three layers of classroom evidence: overall descriptive score change, variation across intact classes, and component-level vulnerability among lower-grade profiles. The overall gain of +6.47 suggests a positive descriptive change from MID-TERM TEST to FINAL TEST. The class-level results show that all classes recorded positive gains, but with different gain sizes and pass rates. The grade distribution shows broad course-level attainment, with 148 of 171 students receiving A or B+. The lower-grade profiles show that this broad attainment coexisted with C/C+, D, and E outcomes linked to weak or missing assessment components.

These patterns support the diagnostic purpose of the study. Institutional assessment records do more than document final grades; they can help lecturers understand how course-based integrated ESP literacy performance is distributed across classes and student profiles. The findings also reinforce the need for methodological caution. Since the study did not use a control group, inferential statistics, classroom observation, or interview data, the results should be interpreted as descriptive assessment evidence associated with the collaborative ESP reading implementation period. The value of the study lies in showing how authentic course records can support pedagogical monitoring, not in proving the causal effectiveness of collaborative learning.

CONCLUSIONS

This study examined assessment patterns during collaborative ESP reading in four Bahasa Inggris 2 classes. Institutional records showed that the mean increased from 68.71 at mid-term to 75.18 at final assessment, a difference of 6.47 points or 9.42%. All classes recorded higher final means, ranging from 3.51 to 9.03 points. The pass rate was 93.57%, with 148 of 171 students receiving A or B+. However, lower-grade profiles revealed zero, low, or imbalanced assessment components. These findings describe performance patterns rather than causal effects. The study contributes a diagnostic perspective by showing how institutional records reveal mean differences, class variation, grade concentration, and lower-grade profiles. Nevertheless, limitations include the one-group design, non-equivalent assessments, absence of inferential tests and effect estimates, and lack of classroom observations or interviews. Future research should use comparable assessments, comparison groups, implementation documentation, and additional qualitative evidence. Thus, institutional records are useful for

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

monitoring and improving ESP instruction but cannot independently prove collaborative-reading effectiveness.

ACKNOWLEDGEMENTS

The author would like to express sincere appreciation to the Management Study Program, Universitas Bumigora, for granting permission to conduct this classroom-based study and to use anonymized institutional assessment records. The author also thanks the students of the Bahasa Inggris 2 course for their participation in the learning process. This study received no specific funding from any public, commercial, or non-profit funding agency.

DECLARATION

Author Contributions: Erwin Suhendra was responsible for the study's conception, examination of assessment records, and preparation of the manuscript. Lalu Arwan Darmajaya contributed to the interpretation of classroom assessment practices, evaluation of the findings, and academic revision. Both authors agreed on the final article.

Conflict of Interest: No circumstances involving competing professional or personal interests are known to have affected this study.

Funding: The authors received no dedicated funding for the research or publication of this article.

Ethical Approval: The study utilized existing course assessment records. Any required institutional authorization and safeguards for student data should be observed in accordance with applicable regulations.

Informed Consent: Consent requirements were addressed in accordance with institutional policies governing the use of student assessment records.

Use of Generative AI: The authors were responsible for the research design, examination of assessment patterns, and interpretation of results. AI-based tools, if used, provided supplementary assistance in preparing the written presentation.

REFERENCES

Aldossary, K. (2024). Evaluating the effectiveness of collaborative strategic reading on EFL female university students' reading comprehension. *Forum for Linguistic Studies*, *6*(6), 785–799. https://doi.org/10.30564/fls.v6i6.7544

Bin Dahmash, N. (2025). "We divided the questions evenly among ourselves to answer them properly": Collaborative learning in EFL classroom status quo. Arab World English Journal (AWEJ), *16* https://doi.org/10.24093/awej/vol16no3.23

Butarbutar, R., Ruing, F. H., Basri, N., Tuharea, V. U., & Leba, S. M. R. (2023). Unpacking online collaborative learning in teaching EFL speaking: Insights from three rural area case studies. T*he Qualitative Report,* 28*(*12), 3379–3401. https://doi.org/10.46743/2160-3715/2023.6165

Butros Abed, T. B. (2019). The implementation of jigsaw as a cooperative learning strategy to improve Birzeit University EFL students' reading comprehension. *Ricerche di Pedagogia e Didattica*, *14*(3), 269–304. https://doi.org/10.6092/issn.1970-2221/9430

Chang-Tik, C. (2023). Collaborative spaces: Investigating the relationships between students' group-based learning and lecturers' approaches. Educational Research, 65(4), 409–427 https://doi.org/10.1080/00131881.2023.2256747

Chinpakdee, M. (2024). EFL teachers' expectations and students' reading-related difficulties at the university level. LEARN Journal: Langu*age Education and Acquisition Research Network, 17(2), 48–68* https://doi.org/10.70730/ytvo7549

Dai, B. (2024). A study of using ESP teaching model to promote reading comprehension skills in English specialized areas. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns-2024-2500

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

Duong, T. M., & Le Truong, M. (2022). Task-based language teaching: A possible remedy for Vietnamese EFL students' ESP reading comp*rehension at a vocational college*. H*umanities, Arts and Social Sciences Studies*, 22(2), 291–306.

Fan, T., & Yan, X. (2025). EFL development of micro reading abilities: The impact of three intervention approaches informed by diagnostic language assessment. L*anguage Assessment Quarterly,* 22(2), 164–192. https://doi.org/10.1080/15434303.2025.2497819

Gaffas, Z. M. (2024). Using social media platform X to enhance student medical English learning: An attempt based on design-based research (DBR) in a medical English for specific purposes (ESP) course. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1). https://doi.org/10.1186/s40862-024-00281-1

Han, Z., Zhou, Z., Huang, C., Wang, X., Chen, G., & Huang, Q. (2025). Investigating the evolution of effective teams among middle school students: Insights from real-time collaborative learning. Interactive Learning Environments, 1–21 *https://doi.org/10.1080/10494820.2025.2568908*

Hassan, A., Alawawda, M., Alzahrani, F., & Naz, N. (2023). Developing an ESP-based language learning environment to help students improve critical thinking skills in written output. I*nformation Sciences Letters,* 12(4), 1131–1140. https://doi.org/10.18576/isl/120431

Hatte, P. R., Patil, S. D., Malge, A. M., & Pande, S. (2025). Innovative approach of learning the quality management course contents through collaborative learning in the classroom. J*ournal of Engineering Education Transformations,* 39(2), 104–110. https://doi.org/10.16920/jeet/2025/v39i2/25145

Hirvela, A. (2012). ESP and Reading. In The Handbook of E*nglish for Specific Purposes (pp. 77–94) https://doi.org/10.1002/9781118339855.ch4*

Hirvela, A. (2025). ESP and Reading. In *The Handbook of English for Specific Purposes: Second Edition* (pp. 73–88). Wiley. https://doi.org/10.1002/9781119985068.ch4

Hirvela, A. (2026). Academic reading for writing. In The Routledge H*andbook of English for Academic Purposes (pp. 155–167). Routledge https://doi.org/10.4324/9781003493358-15*

Hyland, K. (2019). English for specific purposes: Some influences and impacts. In *Springer International Handbooks of Education: Part F1628* (pp. 337–353). Springer Nature. https://doi.org/10.1007/978-3-030-02899-2_19

Jin, T., Liu, X., & Lei, J. (2020). Developing an effective three-stage teaching method for collaborative academic reading: Evidence from Chinese first-year college students. Journal of English for Academic P*urposes, 45, 100853 https://doi.org/10.1016/j.jeap.2020.100853*

Khampool, W., & Chumworatayee, T. (2023). Collaborative strategic reading (CSR) instruction and its effects on thai EAP university learners' reading comprehension across two proficiency levels. LE*ARN journal: Language education and acquisition research network,* 16(2), 8–24. https://www.scopus.com/pages/publications/85169610874

Kinnear, P. (2018). English for specific purposes: Engineers learning how to mean in English. In *The Routledge Handbook of Sociocultural Theory and Second Language Development* (pp. 357–377). Taylor and Francis. https://doi.org/10.4324/9781315624747

Liu, X., Gu, M. M., & Jin, T. (2024). Strategy use in collaborative academic reading: Understanding how undergraduate students co-construct comprehension of academic texts. L*anguage Teaching Research,* 28(4), 1391–1411.

Shaik, M. G. (2025). Pedagogical approaches to collaborative learning: An analytical study. In M. K. R. Vennapusa & A. Shaik (Eds.), A*IP Conference Proceedings* (Vol. 3237, Issue 1). American Institute of Physics. https://doi.org/10.1063/5.0248162

Shen, M.-Y. (2013). Toward an understanding of technical university EFL learners' academic reading difficulties, strategies, and needs. *Electronic Journal of Foreign Language Teaching*, *10*(1).

Stoller, F. L., & Robinson, M. S. (2018). Innovative ESP teaching practices and materials development. In Ke*y Issues in English for Specific Purposes in Higher Education* (pp. 29–49). Springer.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Erwin Suhendra & Lalu Arwan Darmajaya.

Suparto, W. P., Abduh, A., & Basri, M. (2025). Integrating collaborative strategic reading with digital platforms creates a hybrid model for modern English language pedagogy. T*heory and Practice in Language Studies*, *15*(9), 3009–3019. https://doi.org/10.17507/tpls.1509.23

Synekop, O., & Lytovchenko, I. (2024). Integrated writing in differentiated ESP course at technical university. LL*T journal: Journal on language and language teaching*, *27*(2), 599–612. https://doi.org/10.24071/llt.v27i2.8003

Vasconcelos, S. V. (2023). Learning while teaching: Harnessing the potential of peer-to-peer learning to enhance language learning. In *Fostering Pedagogy Through Micro and Adaptive Learning in Higher Education: Trends, Tools, and Applications* (pp. 93–110). IGI Global. https://doi.org/10.4018/978-1-6684-8656-6.ch005

Volet, S., Vauras, M., Salo, A.-E., & Khosa, D. (2017). Individual contributions in student-led collaborative learning: Insights from two analytical approaches to explain the quality of group outcome. L*earning and Individual Differences*, *53*, 79–92.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.