

Tracing Input Use Across Integrated EFL Assessment: Relationships Between Receptive and Productive Performance

Hira Maryana^{*1}, Fazri Nur Yusuf¹, Lulu Laela Amalia¹

¹ Universitas Pendidikan Indonesia, Bandung, Indonesia

Received: May 7, 2026 | Revised: July 21, 2026 | Accepted: August 10, 2026

Correspondence: hiramaryana24@upi.edu

Abstract

Although integrated-skills assessment more closely reflects real-world communication than discrete-skill testing, limited evidence explains how receptive and productive performances are related in Indonesian secondary EFL classrooms. This convergent mixed-methods study examined associations between listening and reading performance and writing and speaking performance and explored how learners used input during integrated assessment. Participants were 128 tenth-grade students from four intact classes at an Indonesian senior high school. They completed listening-writing, reading-writing, and reading-writing-speaking tasks, followed by reflective questionnaires; classroom observations provided additional qualitative evidence. As receptive and productive scores were non-normally distributed, Spearman's rho was used, while qualitative data were thematically analyzed and merged with the correlations. Overall receptive and productive performance showed a very strong positive association ($\rho = .949$, $p < .001$). All skill-pair correlations were significant: listening-writing ($\rho = .658$), listening-speaking ($\rho = .707$), reading-writing ($\rho = .836$), and reading-speaking ($\rho = .800$), all $p < .001$. Qualitative findings explained these patterns: students reused source ideas, vocabulary, and organizational cues in written and oral responses, treated listening input as pronunciation and dialogue models, and used notes to retain and transform information. Reading was perceived as particularly useful for writing, whereas listening was more useful for speaking. These findings indicate interdependence, not causality, between receptive and productive performance. They show that integrated tasks can provide ecologically meaningful evidence of learners' comprehension-to-production processes and support their use in secondary EFL assessment. Future validation research should separate associations generated by shared task content from those reflecting broader language proficiency.

Keywords: EFL assessment; integrated skills; productive skills; receptive skills; secondary education; mixed methods

1. Introduction

Language assessment has conventionally operationalized listening, reading, writing, and speaking as related but distinguishable domains, enabling standardized administration, skill-specific interpretation, and diagnostic reporting (Choi & Papageorgiou, 2020; Dunn & McCray, 2020; Sawaki & Sinharay, 2018). Although useful for identifying learners' strengths and weaknesses in particular skills, this compartmentalization captures only part of communicative language ability because situated language use frequently requires comprehension and production to operate together within goal-directed activity (Banerjee, 2019; Frost et al., 2020; Kunnan et al., 2022). To capture this coordination, integrated EFL assessment employs tasks in which learners process written, oral, or multimodal sources and subsequently use the information to produce a written or spoken response. Common configurations include reading-to-write, listening-to-write, reading-to-speak, listening-to-speak, and tasks combining multiple input modes before production (Kormos et al., 2020; Shin & Ewert, 2015; Wei & Zheng, 2017; Westbrook, 2023). Their distinctive contribution lies in requiring learners not only to comprehend and produce language but also to select relevant content, retain key information, and reshape it for a new communicative purpose. The distinction between discrete and integrated assessment is therefore one of intended interpretation rather than relative superiority. Discrete-skill tasks remain valuable when the intended inference concerns a

specific ability or subskill (Heidari Vincheh et al., 2024; Kim, 2015; Min & He, 2022; Mirzaei et al., 2020), whereas integrated tasks provide a basis for examining how receptive understanding, source use, and productive performance are coordinated within the same communicative event (Harsch et al., 2024; Lestari & Brunfaut, 2023; O'Grady & Taşkesen, 2022).

This broader interpretive purpose is especially evident in source-mediated communication, where learners must consult supplied material, determine its relevance, preserve its intended meaning, and reorganize information according to disciplinary, rhetorical, or interpersonal purposes (Leijten et al., 2019; Neumann et al., 2019; Wette, 2018). Because the resulting response depends partly on the source material, integrated tasks can reveal how learners connect comprehension and production by identifying important content, representing source ideas accurately, synthesizing information, constructing an appropriate organization, and selectively drawing on source language (Chuang & Yan, 2023; Gebril & Plakans, 2016; Kyle, 2020; Uludag et al., 2019). However, the expanded evidence provided by these tasks also introduces greater construct complexity, since performance cannot be interpreted as a pure measure of either the receptive or productive skill involved. Studies have associated integrated-task performance with source comprehension, planning, working-memory capacity, anxiety, note use, and the strategies employed to convert input into output (Güvendir & Uzun, 2023; Huang & Hung, 2018; Payant et al., 2019; Siegel, 2024). At the production stage, performance is further shaped by lexical sophistication, syntactic complexity, self-efficacy, and self-regulatory strategy use (Golparvar & Khafi, 2021; Jiang et al., 2025; Kyle & Crossley, 2016; Zhao et al., 2023). This multidimensionality creates a validity and diagnostic challenge because a weak written or spoken response may arise from inadequate source comprehension, unsuccessful content selection, ineffective note-taking or planning, limited productive language resources, or difficulty formulating information that was otherwise understood. Defensible interpretation consequently requires careful control of source and task characteristics, together with analytic criteria that distinguish content selection, source integration, response organization, and productive language quality (Homayounzadeh et al., 2019; Ismayilli Karakoç et al., 2025; Ohta et al., 2018; Wang et al., 2017). Addressing this interpretive challenge first requires a clearer account of how receptive and productive abilities are related.

At the ability level, this relationship is best understood as structured interdependence. Psychometric and literacy research identifies a shared dimension of general language proficiency while also demonstrating meaningful skill- and task-specific variance, indicating that comprehension and production are connected but neither equivalent nor interchangeable (In'nami et al., 2016; Nakamura, 2022; Schoonen, 2019; Yan et al., 2019). Within this shared system, listening and reading provide access to source ideas, factual information, vocabulary, grammatical patterns, and discourse structures that may be reused or adapted during subsequent production, although learners' uptake of these resources is selective rather than automatic (Wang et al., 2022; Ye & Ren, 2019; Zhu et al., 2016; Zhu et al., 2021). Writing and speaking, however, impose additional demands because learners must retrieve words, formulate sentences, organize ideas, monitor emerging output, and, in oral production, articulate language fluently under real-time constraints (Michel et al., 2020; Révész et al., 2019; Suzuki & Kormos, 2023; Vasylets & Marín, 2021). Comprehension may therefore facilitate production without guaranteeing it, as learners who understand source content may still struggle to retrieve appropriate vocabulary, organize their responses, or express information accurately and fluently (Choi, 2017; Uchihara & Clenton, 2023; Uchihara & Saito, 2019; Zhong, 2018). The strength of these relationships may also vary by modality. Reading and writing share visually available orthographic and discourse resources, whereas listening and speaking depend more strongly on phonological processing and spoken lexical access (Cheng & Matthews, 2018; Saito et al., 2025; Uchihara & Clenton, 2023). Nevertheless, channel congruence is not essential because learners can transform spoken material into writing and reconstruct written information for oral production, making cross-modal transfer central to many integrated tasks (Coumel et al., 2023; Kormos et al., 2022). Relationships among receptive and productive scores should therefore be interpreted as evidence of interdependence rather than proof that receptive ability directly causes productive performance (Liao et al., 2021; Ye & Ren, 2023). This distinction

shifts attention from whether the skills are associated to the processes through which comprehended information becomes a resource for production.

That mediating process is captured by the construct of input use, which encompasses the cognitive, strategic, and textual operations through which learners select, retain, interpret, reorganize, and incorporate information from listening or reading into subsequent written or spoken responses (Cumming et al., 2016; Du & List, 2020; Machili et al., 2020). Input use should not be equated with mechanical reproduction because engagement with source material can occur at different levels of transformation. At the more direct end of the continuum, learners may reproduce factual details, reuse source vocabulary or grammatical patterns, and retain the source's original organization. More generative input use involves paraphrasing ideas, synthesizing information, connecting source content with prior knowledge, restructuring material for a new audience or purpose, and expressing it through independently formulated language (Du & List, 2020; Nelson & King, 2023; Vandermeulen et al., 2020). Viewing input use as a continuum also prevents source integration from being judged solely by the quantity of borrowed ideas or degree of lexical overlap. Source incorporation becomes meaningful when selected information is relevant, accurately represented, appropriately transformed, coherently integrated, and responsive to the communicative demands of the task (Chuang, 2025; Sawaki et al., 2024; Xie, 2023). Extensive borrowing may indicate successful content coverage, limited productive resources, or excessive source dependence, whereas extensive transformation may demonstrate deeper synthesis but also introduce distortion or omission. Input use must therefore be evaluated according to the quality and function of source incorporation rather than its frequency alone.

How learners realize these different forms of input use depends partly on the affordances and processing demands of the input modality. Written sources remain visually available for repeated inspection of lexical, grammatical, and organizational features, whereas spoken sources unfold temporally and expose learners to pronunciation, stress, rhythm, intonation, and spoken-discourse cues while placing greater demands on attention and memory (Arias-Hermoso et al., 2024; Cai & Huang, 2023; Pu & Chang, 2025; Saito & Saito, 2017). These contrasting affordances make note-taking particularly important as a bridge between input and output because it can direct attention toward important content, create an external record of transient information, reduce reliance on unaided memory, and provide material for subsequent planning and reconstruction (Jin & Webb, 2024; Siegel, 2018, 2020; Yeager et al., 2024). Because no single indicator can fully represent this process, input use can be traced through complementary evidence, including learners' notes, source-consultation patterns, idea and lexical uptake, paraphrasing, response organization, strategy reports, and process data recorded during task completion (Delgado-Osorio et al., 2023, 2025; Lu et al., 2025; Ranalli et al., 2018). Bringing these indicators together can clarify how comprehended information becomes a resource for production while recognizing that learners with comparable receptive performance may use the same input differently because of variations in proficiency, topical knowledge, strategic behavior, executive functioning, and source-integration ability (Crossley & Kim, 2019; Huang et al., 2016; Liao et al., 2025; Zhu et al., 2025). A process-oriented account of input use thus provides a conceptual and analytical bridge between statistical relationships among skill scores and explanations of how those relationships emerge during task performance.

Despite this potential, existing research rarely brings these two levels of evidence together. Many studies concentrate on a single task configuration, particularly reading-to-writing or listening-to-writing, rather than systematically examining how both listening and reading relate to writing and speaking. Score-based studies have established relationships among receptive and productive skills but provide limited insight into the processes underlying those relationships. Conversely, process-oriented studies of source use, note-taking, and learner strategies seldom connect such evidence with composite and skill-specific performance patterns. This fragmentation is particularly apparent in classroom-based integrated assessment at the secondary-school level, especially in Indonesian EFL contexts. To address these gaps, the present convergent mixed-methods study examines the overall relationship between receptive and productive performance, compares the specific associations of listening and reading with writing and speaking, and explores how learners use oral and written input when constructing their responses. Its novelty lies in integrating composite and skill-

pair correlations with qualitative evidence of learners' input-use practices across multiple receptive-productive configurations. By connecting performance patterns with learners' reuse of source ideas and vocabulary, attention to organizational and pronunciation models, and note-mediated retention and transformation of information, the study moves beyond determining whether receptive and productive performances are associated to explain how their interdependence is manifested during integrated assessment. It therefore offers a process-informed and contextually grounded account of input use among Indonesian secondary EFL learners without interpreting the identified relationships as causal effects.

2. Method

2.1 Research Design

This study employed a convergent mixed-methods design in which quantitative and qualitative data were collected during the same integrated-skills assessment phase, analyzed separately, and subsequently merged for interpretation. The defining methodological feature of such a design is not simply the inclusion of two forms of data, but their intentional integration to produce conclusions that neither strand could support independently (Hirose & Creswell, 2023; Johnson et al., 2007). The quantitative strand used students' receptive- and productive-skill scores to examine whether stronger performance in listening and reading was associated with stronger performance in writing and speaking. The qualitative strand drew on classroom observations and reflective questionnaires completed after the integrated tasks to investigate how students used listening and reading input when constructing written and spoken responses. Following the separate analyses, the two strands were integrated through a joint display that aligned the correlation results with relevant qualitative themes. Joint displays can function as analytic devices for identifying agreement, complementarity, inconsistency, and expansion across quantitative and qualitative evidence rather than merely as formats for presenting results (Haynes-Brown & Feters, 2021). The integrated interpretation produced meta-inferences about how the observed input-use practices helped explain the statistical patterns (Younas et al., 2023). The design therefore enabled the study to estimate the strength and direction of receptive-productive relationships while providing process-oriented evidence with which to interpret those relationships in an Indonesian senior-high-school EFL assessment context. Because the quantitative strand was correlational, the qualitative findings were used to contextualize the statistical associations rather than to establish causal mechanisms (Rohrer, 2018).

2.2 Context and Participants

The study was conducted at an Indonesian senior high school where English was taught as a foreign language within the national curriculum, which emphasizes student-centered instruction, contextualized assessment, and the integration of language skills (Zein et al., 2020). The research site was purposively selected because English instruction in Grade 10 incorporated multiple language skills and provided an appropriate setting for implementing integrated assessment tasks. Participants were 128 tenth-grade students from four intact classes, and their existing class membership was retained throughout the study. Purposeful sampling is appropriate when cases are selected for their relevance to the research questions, but it does not provide the same basis for population generalization as probability sampling (Teddlie & Yu, 2007). Although the students had previously participated in communicative classroom activities, they had limited experience completing assessment tasks that systematically combined receptive and productive skills. This context was therefore relevant for examining relationships among listening, reading, writing, and speaking performance and investigating how students used source input across different task configurations.

Permission to conduct the study was obtained from the school, and parental or guardian consent and student assent were secured before data collection. Participation in the research and the use of students' assessment data were voluntary, and all data were anonymized before analysis. Because the school and intact classes were selected purposively, the sample was intended to provide a relevant and information-rich context for addressing the research questions rather than statistical representation of all Indonesian senior-high-school students (Teddlie & Yu, 2007).

2.3 Research Instruments

The primary instrument was a summative integrated-skills assessment designed to measure students' receptive and productive English performance. Integrated language-assessment tasks require the productive response to depend meaningfully on comprehension and use of the supplied input; accordingly, both the source-processing demands and the expected written or spoken response need to be represented in the task design (Frost et al., 2020; Rukthong & Brunfaut, 2020). The assessment comprised three task configurations. First, in the listening-writing task, students listened to short and extended dialogues and completed matching, multiple-choice, and constructed-response items. Second, the reading-writing task required students to comprehend short written passages and respond to completion and extended-response items. Third, in the reading-writing-speaking task, students read an email, composed an appropriate written reply, and participated in a paired conversation based on the same topic. The tasks were developed with attention to communicative purpose, contextual relevance, and meaning-focused language use.

Although the tasks combined multiple skills, each skill component was scored separately to support skill-specific and cross-skill analyses. Because the interpretability of component scores depends on their psychometric quality, component-level reliability was examined before the scores were used in cross-skill comparisons (Sawaki & Sinharay, 2018). The listening and reading components generated comprehension scores and were classified as receptive performance, whereas the designated writing and speaking components generated production scores and were classified as productive performance. Receptive responses were evaluated according to comprehension accuracy, while written and spoken performances were assessed using analytic rubrics. Analytic scales make dimension-level judgments explicit and can support more consistent evaluation of source-dependent performance when the scoring criteria are clearly operationalized (Lestari & Brunfaut, 2023). The writing rubric addressed content relevance, organization, vocabulary, and grammar, while the speaking rubric evaluated content relevance, vocabulary, grammar, and pronunciation.

Evidence based on test content was obtained through expert review. Four experienced EFL teachers examined the alignment of the assessment tasks and individual items with the learning objectives specified in the national curriculum. Expert judgment is a recognized source of content-related validity evidence when reviewers evaluate the relevance and representativeness of test content in relation to a clearly defined domain (Downing & Haladyna, 1997; Haynes et al., 1995). A pilot administration was subsequently conducted with a smaller group of students to evaluate item clarity, task instructions, and the adequacy of the allocated completion time. The expert comments and pilot evidence were then used to revise and finalize the assessment before the main administration, consistent with the use of quality-assurance procedures during test development (Downing & Haladyna, 1997).

Two supporting instruments were used to collect qualitative data. First, a reflective questionnaire invited students to describe their experiences, difficulties, and strategies when completing the integrated tasks, particularly how they drew on listening and reading input when writing or speaking. Second, a structured classroom-observation checklist was used during task administration to record students' engagement and observable task-interaction patterns. The assessment scores, reflective responses, and observation records supplied complementary evidence about students' performance and input-use practices. Their combination enabled data-source triangulation aimed at increasing analytic breadth and examining convergence as well as inconsistency across evidence sources (Fielding, 2012).

2.4 Data Collection

Data were collected over three weeks at the end of the first semester as part of the school's summative-assessment period. Students completed the three task configurations—listening-writing, reading-writing, and reading-writing-speaking—sequentially during regular class hours. All assessment components were completed individually except for the speaking activity, which required students to interact in pairs. Classroom observations were conducted concurrently with task administration to document students' engagement and observable behavior during the integrated activities. After completing the assessment tasks,

students completed the reflective questionnaire describing their experiences, difficulties, and strategies for using the source material. This procedure generated three complementary forms of data: performance scores from the integrated assessment, students' self-reported reflections, and observational evidence recorded during task completion.

2.5 Data Analysis

The quantitative data were analyzed using IBM SPSS Statistics Version 27. Before the main analyses, the data were screened for completeness and examined for distributional characteristics. The internal consistency of the multi-item listening and reading components was summarized using coefficient alpha, while corrected item-total correlations were examined as item-level diagnostic evidence. Because alpha is affected by test length and restrictive measurement assumptions, it was interpreted as one index of score consistency rather than as proof of unidimensionality or sufficient evidence of reliability by itself (McNeish, 2018; Sijtsma, 2009). Writing and speaking performances were independently evaluated by trained raters using the analytic rubrics, and inter-rater reliability was estimated from the scores assigned to the same performances. The reliability coefficient was selected in accordance with the scale of the scores and the structure of the rater design, as recommended for the defensible assessment of rater agreement (Hallgren, 2012).

Because preliminary diagnostics indicated that the score distributions and bivariate relationships did not adequately satisfy the requirements for Pearson's product-moment correlation, Spearman's rank-order correlation coefficient was used. Spearman's coefficient estimates the strength of a monotonic association and is generally less sensitive than Pearson's coefficient to heavy-tailed distributions and influential outliers (de Winter et al., 2016). The correlational analysis was conducted at two levels. First, the relationship between the composite receptive score, comprising listening and reading, and the composite productive score, comprising writing and speaking, was examined. Second, separate correlations were calculated for listening-writing, listening-speaking, reading-writing, and reading-speaking performance. Statistical significance and the magnitude and direction of each coefficient were considered when interpreting the results. These analyses were used exclusively to identify associations between receptive and productive performance and were not used to claim directionality, explained contribution, or causal influence (Rohrer, 2018).

The open-ended reflective-questionnaire responses and qualitative observation records were analyzed thematically. The analysis involved repeated familiarization with the data, initial coding of relevant segments, organization of related codes into candidate themes, and iterative review, definition, and naming of the resulting themes (Braun & Clarke, 2006; Nowell et al., 2017). The analysis focused on students' experiences, difficulties, and strategies when obtaining information from listening and reading sources and using that information in written and spoken responses. Themes were compared across the reflective questionnaires and observation records to identify recurring, complementary, and contrasting patterns. Data-source triangulation was used to broaden and interrogate the qualitative interpretation, not to treat agreement among sources as automatic proof of validity (Fielding, 2012).

In the final integration stage, the quantitative and qualitative results were brought together in a joint display. Each composite or skill-pair correlation was aligned with the qualitative themes and supporting evidence relevant to that relationship. The display was examined for convergence, complementarity, expansion, and divergence across the two strands, using integration as an analytic process rather than only a reporting technique (Haynes-Brown & Feters, 2021; Hirose & Creswell, 2023). The resulting meta-inferences provided a more comprehensive account of how receptive and productive performances were related and how students used listening and reading input during integrated assessment (Younas et al., 2023). These meta-inferences remained relational and interpretive; the qualitative evidence was not treated as proof that receptive performance caused productive performance.

3. Findings

3.1 Distribution of Receptive- and Productive-Performance Scores

The distributions of the composite receptive- and productive-performance scores were examined before conducting the correlational analysis. The results of the Kolmogorov–Smirnov and Shapiro–Wilk tests are presented in Table 1.

Table 1

Tests of normality for receptive- and productive-performance scores

Variable	K-S statistic	df	p	S-W statistic	df	p
Receptive performance	.119	128	< .001	.943	128	< .001
Productive performance	.150	128	< .001	.932	128	< .001

Both normality tests were statistically significant for the receptive- and productive-performance scores. The receptive score produced a Kolmogorov–Smirnov statistic of .119 and a Shapiro–Wilk statistic of .943, with $p < .001$ for both tests. The productive score produced corresponding statistics of .150 and .932, also with $p < .001$. These results indicated that both score distributions departed significantly from normality. Spearman’s rank-order correlation was therefore used to examine the relationships between receptive and productive performance.

3.2 Overall and Skill-Specific Correlations

The correlational analysis was conducted at two levels. The first analysis examined the relationship between the composite receptive score, comprising listening and reading, and the composite productive score, comprising writing and speaking. The second analysis examined the four individual skill pairings: listening–writing, listening–speaking, reading–writing, and reading–speaking. The results are consolidated in Table 2.

Table 2

Spearman correlations between receptive and productive performance

Analysis level	Relationship	Spearman’s ρ	p	N
Composite	Receptive–productive	.949	< .001	128
Skill-specific	Listening–writing	.658	< .001	128
Skill-specific	Listening–speaking	.707	< .001	128
Skill-specific	Reading–writing	.836	< .001	128
Skill-specific	Reading–speaking	.800	< .001	128

At the composite level, receptive and productive performance demonstrated a very strong positive relationship, $\rho = .949$, $p < .001$. Students with higher combined listening and reading scores therefore tended to have higher combined writing and speaking scores. This result represents a close correspondence in students’ rank ordering across the two composite measures; it does not demonstrate that receptive performance caused or improved productive performance.

All four skill-specific relationships were also positive and statistically significant. The largest coefficient was found between reading and writing, $\rho = .836$, $p < .001$, followed by reading and speaking, $\rho = .800$, $p < .001$. Listening was positively associated with speaking, $\rho = .707$, $p < .001$, and writing, $\rho = .658$, $p < .001$. Numerically, reading was more strongly associated with both productive skills than listening was, while listening demonstrated a larger coefficient with speaking than with writing. However, because no formal statistical comparison of the correlation coefficients was conducted, these numerical differences should not be interpreted as statistically significant differences between relationships.

3.3 Descriptive Questionnaire Patterns

The closed-response questionnaire items asked students which receptive input they considered helpful when completing the writing and speaking tasks. The supplied percentages are presented in Table 3.

Table 3

Receptive input identified by students as helpful for production

Productive task	Listening input (%)	Reading input (%)
Writing	30	70
Speaking	68	38*

For writing, reading was identified more frequently than listening as a helpful source of input. For speaking, listening was identified more frequently than reading. However, the exact speaking distribution cannot be finalized until the percentage discrepancy shown in Table 3 has been resolved.

3.4 Qualitative Findings

The qualitative data from students' reflective questionnaires and structured observation records were analyzed separately from the statistical data. Initial codes were grouped into recurring patterns describing how students used listening and reading input when producing written and spoken responses. Four themes emerged: (1) reading input as lexical and content support for writing, (2) listening input as a pronunciation and dialogue model for speaking, (3) note-taking as a bridge between input and production, and (4) selective reuse and elaboration of source information. In the excerpts below, SRQ refers to a student reflective-questionnaire response and OBS refers to an observation record. Each excerpt is followed by its evidence-source identifier and analytic code.

3.4.1 Reading Input as Lexical and Content Support for Writing

Students frequently described the reading passages as sources of vocabulary, ideas, and textual examples for their written responses. One student stated, *"I use vocabulary from the text in my writing"* (SRQ-01, code: lexical reuse). Another explained, *"It is easier to write because the text give me example"* (SRQ-02, code: textual modelling). These excerpts indicate that students consulted the written input for language that could be incorporated into their responses and for examples of how relevant information could be presented. Other reflective responses similarly referred to obtaining useful words and content from the passages before beginning the writing task. The reading material was therefore used both as a source of information and as a linguistic resource, with students selecting vocabulary, identifying relevant ideas, and drawing on textual examples when constructing their responses. This pattern corresponded with the observation records, which showed students repeatedly returning to the reading passages while planning and composing their written work (OBS-01, code: source consultation).

3.4.2 Listening Input as a Pronunciation and Dialogue Model for Speaking

Students' reflections on the speaking task focused primarily on the oral-language features available in the listening input. One student reported, *"I learned the pronunciation from the listening"* (SRQ-03, code: pronunciation modelling), while another stated, *"Listening helped me with example of dialogues"* (SRQ-04, code: dialogue modelling). These responses show that students attended not only to the meaning of the recorded material but also to how words were pronounced and how conversational exchanges were organized. The listening input consequently provided models of pronunciation, sentence construction, and turn-taking that students could draw on when preparing for the paired conversation. Observation records also indicated that some students rehearsed words or short expressions encountered in the listening material before performing the speaking task (OBS-02, code: oral rehearsal). Listening was therefore used as an immediate spoken-language model rather than solely as a source of factual information.

3.4.3 Note-Taking as a Bridge Between Input and Production

Note-taking emerged as a recurring behavior connecting the receptive and productive stages of the assessment. The observation record noted that *“students recorded keywords and important content points while completing the listening and reading tasks”* (OBS-03, code: input recording). A subsequent entry documented that *“students referred to their notes when selecting information and organizing their written and spoken responses”* (OBS-04, code: note reuse). These coded excerpts demonstrate that students created an external record of information obtained during comprehension and later used it to support production. The practice was particularly evident during the listening task because the oral input was no longer available after the recording ended. Students’ notes preserved selected words and ideas that could subsequently be retrieved during writing and speaking. Note-taking therefore functioned as an observable input–output bridge through which students retained, selected, and transferred information across different stages of the integrated assessment.

3.4.4 Selective Reuse and Elaboration of Source Information

The observation records showed that students varied in how directly they incorporated source material into their responses. One coded entry indicated that *“some students reused vocabulary, sentence patterns, and content ideas directly from the source materials”* (OBS-05, code: direct uptake). Another recorded that *“other students reorganized the source ideas and added further details when producing their written and spoken responses”* (OBS-06, codes: reformulation and elaboration). Thus, input use did not consist exclusively of reproducing information in its original form. Some students relied on direct lexical or structural borrowing, whereas others selected relevant content, expressed it differently, reorganized its presentation, or connected it with additional details. These patterns represent different degrees of transformation, ranging from close source dependence to more independently formulated production.

The four themes show that students used receptive input through several interconnected practices: drawing vocabulary and ideas from reading passages, using listening material as a model for pronunciation and dialogue, recording information in notes for later retrieval, and either reproducing or transforming source content during production. These findings describe the ways in which input was reportedly and observably used during the integrated assessment; they do not establish that these practices caused the students’ productive-performance scores.

4. Discussion

This study examined the relationship between receptive and productive performance and explored how Indonesian secondary-school EFL learners used listening and reading input when constructing written and spoken responses. The quantitative findings revealed a very strong positive relationship between the composite receptive and productive scores, together with strong positive relationships across all four skill pairings. Reading demonstrated the numerically strongest relationships with writing and speaking, whereas listening was more strongly associated with speaking than with writing. The qualitative findings complemented these patterns by showing that students used reading as a source of vocabulary and content, listening as a model for pronunciation and dialogue, notes as an external record of source information, and both oral and written input for direct reuse and more elaborated reconstruction. Taken together, the findings portray receptive and productive abilities as coordinated but distinguishable dimensions of performance and identify input use as an important process connecting them during integrated assessment.

4.1 Receptive–Productive Interdependence in Integrated Assessment

The very strong relationship between the composite receptive and productive scores, Spearman’s $\rho = .949$, $p < .001$, indicates that students’ rank ordering on listening and reading performance closely corresponded with their rank ordering on writing and speaking performance. This result is consistent with accounts of second-language proficiency as comprising a shared general-language dimension alongside meaningful skill-specific variation. Receptive and productive abilities consequently draw on overlapping lexical, grammatical, and discourse knowledge but remain distinguishable because they require different

configurations of that knowledge during task performance (In'nami et al., 2016; Nakamura, 2022; Schoonen, 2019; Yan et al., 2019). In the present assessment, students who comprehended source ideas, vocabulary, and discourse relationships more successfully also tended to produce more successful written and spoken responses because the productive tasks required them to operate on information obtained from the receptive materials.

Nevertheless, the magnitude of the composite correlation should not be interpreted as demonstrating that receptive and productive abilities are equivalent or that comprehension directly determines production. Writing and speaking impose additional demands involving lexical retrieval, grammatical formulation, discourse organization, monitoring, and, particularly in speaking, real-time articulation (Michel et al., 2020; Révész et al., 2019; Suzuki & Kormos, 2023; Vasylets & Marín, 2021). Learners may therefore comprehend source content without possessing sufficient productive resources to express that understanding accurately, coherently, or fluently (Choi, 2017; Uchihara & Saito, 2019; Zhong, 2018). The exceptionally high composite coefficient may also partly reflect score aggregation and the dependence created by using common topics and source materials across receptive and productive components. It is therefore more defensible to interpret the coefficient as evidence of close coordination within the integrated assessment event rather than as evidence that receptive performance caused productive performance.

4.2 Relationships Across Specific Skill Pairings

Among the four skill pairings, reading and writing demonstrated the strongest relationship, $\rho = .836, p < .001$. The qualitative finding that students repeatedly drew vocabulary, ideas, and textual examples from the reading passages provides a process-oriented account of this relationship. Unlike spoken input, written material remained visually available during task completion, enabling students to revisit particular words, verify content, inspect sentence patterns, and reconsider how source ideas could be incorporated into their responses. Research on source-mediated writing similarly shows that reading input can supply content, lexical resources, discourse structures, and rhetorical models, although learners use these resources selectively according to their comprehension, proficiency, and writing goals (Wang et al., 2022; Ye & Ren, 2019; Zhu et al., 2016, 2021).

The students' reported use of vocabulary and examples from the passages also corresponds with research showing that integrated writing requires more than general reading comprehension. Learners must identify relevant content, represent it accurately, organize it according to the response purpose, and express it using appropriate linguistic resources (Chuang & Yan, 2023; Gebril & Plakans, 2016; Kyle, 2020; Uludag et al., 2019). Thus, the strong reading-writing correlation likely reflects both shared language resources and direct task dependency: successful writing partly depended on understanding and using the supplied written material. This interpretation does not establish that reading is inherently more important than listening for writing because the magnitude of the relationship may also have been shaped by the continued availability, content, difficulty, and rhetorical structure of the reading passages.

Reading was also strongly related to speaking, $\rho = .800, p < .001$. This finding demonstrates that receptive-productive coordination was not restricted to skills operating through the same channel. Written input could provide students with ideas, vocabulary, and an organizational basis that they subsequently reformulated for oral interaction. The relationship therefore illustrates cross-modal input use, in which information encoded visually is reconstructed through spoken production (Coumel et al., 2023; Kormos et al., 2022). Its magnitude also challenges an exclusively modality-based explanation in which reading would be expected to relate primarily to writing and listening primarily to speaking. Instead, the usefulness of an input source appears to depend on the informational and linguistic resources it provides and on learners' capacity to transform those resources for a new mode and communicative purpose.

Such transformation is not automatic. When moving from reading to speaking, learners must convert visually available language into orally retrievable representations, organize content rapidly, formulate utterances, and manage interaction with a partner. These additional demands help explain why

comprehension and production remain distinguishable even when their scores are strongly related. The reading material may have reduced the conceptual burden of deciding what to say, but successful speaking still depended on productive vocabulary, grammatical formulation, pronunciation, fluency, and interactional management. The result therefore represents coordinated cross-modal performance rather than direct transfer from reading ability to speaking ability.

Listening was strongly associated with both speaking, $\rho = .707, p < .001$, and writing, $\rho = .658, p < .001$. The numerically stronger listening–speaking relationship corresponds with the students’ descriptions of listening input as a source of pronunciation, sentence patterns, and dialogue models. Listening and speaking share phonological processing and spoken lexical access, allowing learners to draw on the pronunciation, rhythm, and conversational organization encountered in oral input when preparing their own spoken responses (Cheng & Matthews, 2018; Saito et al., 2025; Uchihara & Clenton, 2023). The observed rehearsal of words and expressions further suggests that some students did not attend exclusively to propositional meaning but also treated the listening material as a model of how the target language could be orally realized.

The listening–writing relationship demonstrates another form of cross-modal coordination. To use oral information in writing, students needed to retain relevant content after the recording had unfolded and subsequently reconstruct it through written language. Compared with reading, listening input is transient and cannot normally be inspected repeatedly unless replay is permitted. It places comparatively greater demands on attention, segmentation, short-term retention, and the identification of important information while the input is still unfolding (Arias-Hermoso et al., 2024; Cai & Huang, 2023; Pu & Chang, 2025). These processing demands may help explain why listening demonstrated a numerically smaller relationship with writing than reading did. However, because the correlation coefficients were not formally compared, the ordering should be treated as a descriptive pattern rather than evidence that one relationship was statistically stronger than another.

4.3 Input Use as the Connection Between Comprehension and Production

The qualitative findings clarify that receptive–productive relationships were realized through multiple forms of input use rather than through the automatic transfer of comprehension into production. Students selected vocabulary and content from reading passages, attended to pronunciation and conversational structures in listening materials, recorded information in notes, and subsequently reused or reformulated source information. These practices correspond with a multidimensional conception of input use involving the selection, retention, interpretation, reorganization, and incorporation of source material into a new response (Cumming et al., 2016; Du & List, 2020; Machili et al., 2020). Receptive performance therefore provided access to possible resources for production, while students’ strategic and linguistic decisions determined how those resources were subsequently used.

Note-taking was particularly important because it provided an observable bridge between input reception and later production. Students recorded keywords and content points and subsequently consulted those notes when planning their written and spoken responses. For listening, notes preserved selected information after the source was no longer available; for reading, they helped condense dispersed information into a more manageable basis for response construction. These functions are consistent with evidence that note-taking can direct attention toward important information, externalize memory, and support the planning and reconstruction of source-based responses (Jin & Webb, 2024; Siegel, 2018, 2020; Yeager et al., 2024). Nevertheless, the present data show that students took and reused notes; they do not establish that note-taking itself improved their scores. Establishing such an effect would require systematic analysis of note quality and its relationship with individual performance.

The findings also show that input use occurred along a continuum from relatively direct uptake to more generative transformation. Some students reused source vocabulary, sentence patterns, or content ideas with limited modification, whereas others reorganized information, reformulated ideas, or extended them with additional details. This variation supports the distinction between source reproduction and source

transformation found in research on integrated performance (Nelson & King, 2023; Vandermeulen et al., 2020). Importantly, direct uptake should not automatically be treated as unsuccessful input use, because the accurate retention of key terminology or factual details may be appropriate for particular tasks. Similarly, extensive reformulation does not necessarily indicate superior performance if it changes the source meaning or introduces irrelevant information.

The quality of input use must consequently be judged according to the relevance, accuracy, function, and coherence of the incorporated material rather than the amount of lexical overlap or number of borrowed ideas alone (Chuang, 2025; Sawaki et al., 2024; Xie, 2023). The present findings extend this position across several receptive-productive configurations. Reading supplied durable lexical and informational resources, listening provided both content and oral-language models, and notes supported the retention and reorganization of information across modalities. Input use therefore operated as a flexible process shaped by the characteristics of the source, the demands of the productive task, and the learner's capacity to convert available information into an appropriate response.

The integration of the quantitative and qualitative findings consequently offers a more informative account than either strand could provide independently. The correlations demonstrate that receptive and productive performances covaried, whereas the qualitative themes identify practices through which source information became available for written and spoken production. The themes should not, however, be treated as tested mediators of the statistical relationships. Unless students' coded input-use practices are linked individually with their performance scores, the qualitative findings provide a contextually plausible process account rather than proof that particular practices produced higher achievement. Learners with comparable comprehension scores may use input differently because of variation in prior knowledge, proficiency, strategic behavior, executive functioning, and source-integration ability (Crossley & Kim, 2019; Huang et al., 2016; Liao et al., 2025; Zhu et al., 2025).

4.4 Assessment and Pedagogical Implications

The findings have implications for interpreting integrated EFL assessment. First, an integrated-task score should not be treated as a pure measure of either comprehension or production. A weak response may result from incomplete source comprehension, unsuccessful content selection, ineffective note-taking, limited productive language resources, or difficulty organizing information that was adequately understood. Reporting separate listening, reading, writing, and speaking scores alongside composite results can therefore provide more diagnostically meaningful information. Analytic scoring criteria should also distinguish source-content relevance and accuracy, degree of integration, response organization, and productive language quality (Homayounzadeh et al., 2019; Ismayilli Karakoç et al., 2025; Ohta et al., 2018; Wang et al., 2017).

Second, source and task characteristics require careful control because the availability and modality of input affect how students can use it. Reading passages may be repeatedly consulted, whereas listening materials require immediate processing and retention unless replay is permitted. Note-taking instructions, replay conditions, source length, topic familiarity, response time, and access to the original source should therefore be standardized and reported. Otherwise, correlations among skill scores may reflect differences in task affordances as much as relationships among the intended language abilities.

Pedagogically, the findings suggest that preparation for integrated assessment should develop strategic source use in addition to isolated language skills. Students may benefit from explicit practice in identifying key ideas, recording efficient notes, distinguishing essential from supporting information, paraphrasing without changing meaning, organizing source-based responses, and using spoken input as a model for pronunciation and interaction. These are implications for instruction rather than demonstrated intervention effects, because the present study did not experimentally test whether teaching these strategies improves productive performance.

5. Conclusion

This study examined relationships between receptive and productive performance and explored how Indonesian senior-high-school EFL learners used listening and reading input during integrated assessment. The results revealed a very strong positive association between composite receptive and productive scores ($\rho = .949, p < .001$), together with strong positive associations across all four skill pairings. Reading-writing produced the largest numerical coefficient, followed by reading-speaking, listening-speaking, and listening-writing. This pattern indicates that receptive and productive performances were closely coordinated within the assessment, including across written and spoken modalities. However, the numerical differences among the skill pairings were not formally compared, and the correlations should not be interpreted as evidence that receptive performance caused improvement in productive performance. The qualitative findings clarified how this coordination occurred during task completion. Students used reading passages as sources of vocabulary, content, and textual models, whereas listening materials provided pronunciation, sentence-pattern, and dialogue models for speaking. Note-taking helped students retain and retrieve information across the receptive and productive stages, while source use ranged from direct lexical or idea uptake to reformulation, reorganization, and elaboration. Input use was therefore not a mechanical transfer of comprehended material into production but a selective process involving retention, interpretation, and transformation according to the demands of the response task.

These findings contribute to integrated EFL assessment by showing that productive performance reflects not only writing or speaking ability but also learners' capacity to comprehend, select, preserve, and incorporate source information. Integrated assessments should consequently report skill-specific results alongside composite scores and use scoring criteria that distinguish source comprehension and integration from the linguistic quality of production. Instructionally, learners should be given explicit practice in identifying relevant information, taking effective notes, paraphrasing accurately, organizing source-based responses, and adapting written or spoken input for a different communicative mode. The conclusions remain limited by the single-school sample, correlational design, shared source materials, and reliance on questionnaire and observational evidence. Interpretation of the composite results also depends on establishing component-level reliability for listening and reading and inter-rater reliability for writing and speaking. Future multisite and longitudinal studies should combine individual performance scores with direct analyses of notes, source consultation, lexical uptake, idea integration, and stimulated-recall evidence. Such research could determine more precisely which forms of input use distinguish successful integrated performance and whether explicit input-use instruction contributes to productive-language development over time.

References

- Arias-Hermoso, R., Imaz Agirre, A., & Garro Larrañaga, E. (2024). A comparison between input modalities and languages in source-based multilingual argumentative writing. *Assessing Writing*, 60, 100813. <https://doi.org/10.1016/j.asw.2024.100813>
- Banerjee, H. L. (2019). Investigating the construct of topical knowledge in second language assessment: A scenario-based assessment approach. *Language Assessment Quarterly*, 16(2), 133–160. <https://doi.org/10.1080/15434303.2019.1628237>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Cai, D., & Huang, L. (2023). Effects of input modality on alignment in continuation writing. *Journal of Second Language Writing*, 62, 101060. <https://doi.org/10.1016/j.jslw.2023.101060>
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3–25. <https://doi.org/10.1177/0265532216676851>

- Choi, I., & Papageorgiou, S. (2020). Evaluating subscore uses across multiple levels: A case of reading and listening subscores for young EFL learners. *Language Testing*, 37(2), 254–279. <https://doi.org/10.1177/0265532219879654>
- Choi, Y. H. (2017). Roles of receptive and productive vocabulary knowledge in L2 writing through the mediation of L2 reading ability. *English Teaching*, 72(1), 3–24. <https://doi.org/10.15858/engtea.72.1.201703.3>
- Chuang, P.-L. (2025). Do source use features impact raters' judgment of argumentation? An experimental study. *Language Testing*, 42(1), 48–72. <https://doi.org/10.1177/02655322241263629>
- Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, 60, Article 101003. <https://doi.org/10.1016/j.jslw.2023.101003>
- Coumel, M., Ushioda, E., & Messenger, K. (2023). Second language learning via syntactic priming: Investigating the role of modality, attention, and motivation. *Language Learning*, 73(1), 231–265. <https://doi.org/10.1111/lang.12522>
- Crossley, S. A., & Kim, Y. (2019). Text integration and speaking proficiency: Linguistic, individual differences, and strategy use considerations. *Language Assessment Quarterly*, 16(2), 217–235. <https://doi.org/10.1080/15434303.2019.1628239>
- Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, 23, 47–58. <https://doi.org/10.1016/j.jeap.2016.06.002>
- de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, 21(3), 273–290. <https://doi.org/10.1037/met0000079>
- Delgado-Osorio, X., Hartig, J., Harsch, C., & Koval, V. (2025). Students' behavioral patterns in integrated writing tasks: A sequence analysis approach. *Journal of Educational Psychology*, 117(6), 898–917. <https://doi.org/10.1037/edu0000957>
- Delgado-Osorio, X., Koval, V., Hartig, J., & Harsch, C. (2023). Strategic processing of source text in reading-into-writing tasks: A comparison between summary and argumentative tasks. *Journal of English for Academic Purposes*, 62, 101227. <https://doi.org/10.1016/j.jeap.2023.101227>
- Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10(1), 61–82. https://doi.org/10.1207/s15324818ame1001_4
- Du, H., & List, A. (2020). Researching and writing based on multiple texts. *Learning and Instruction*, 66, 101297. <https://doi.org/10.1016/j.learninstruc.2019.101297>
- Dunn, K. J., & McCray, G. (2020). The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Frontiers in Psychology*, 11, 1357. <https://doi.org/10.3389/fpsyg.2020.01357>
- Fielding, N. G. (2012). Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of Mixed Methods Research*, 6(2), 124–136. <https://doi.org/10.1177/1558689812437101>
- Frost, K., Clothier, J., Huisman, A., & Wigglesworth, G. (2020). Responding to a TOEFL iBT integrated speaking task: Mapping task demands and test takers' use of stimulus content. *Language Testing*, 37(1), 133–155. <https://doi.org/10.1177/0265532219860750>
- Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes*, 24, 78–88. <https://doi.org/10.1016/j.jeap.2016.10.001>

- Golparvar, S. E., & Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assessing Writing*, 47, Article 100504. <https://doi.org/10.1016/j.asw.2020.100504>
- Güvendir, E., & Uzun, K. (2023). L2 writing anxiety, working memory, and task complexity in L2 written performance. *Journal of Second Language Writing*, 60, Article 101016. <https://doi.org/10.1016/j.jslw.2023.101016>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing*, 62, 100894. <https://doi.org/10.1016/j.asw.2024.100894>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Haynes-Brown, T. K., & Feters, M. D. (2021). Using joint display as an analytic process: An illustration using bar graphs joint displays from a mixed methods study of how beliefs shape secondary school teachers' use of technology. *International Journal of Qualitative Methods*, 20, Article 1609406921993286. <https://doi.org/10.1177/1609406921993286>
- Heidari Vincheh, M., Mirzaei, A., & Roohani, A. (2024). A cognitive diagnostic approach to IELTS speaking test: Unveiling the subskills and test-takers' perceptions. *Language Testing in Asia*, 14, Article 42. <https://doi.org/10.1186/s40468-024-00311-2>
- Hirose, M., & Creswell, J. W. (2023). Applying core quality criteria of mixed methods research to an empirical study. *Journal of Mixed Methods Research*, 17(1), 12–28. <https://doi.org/10.1177/15586898221086346>
- Homayounzadeh, M., Saadat, M., & Ahmadi, A. (2019). Investigating the effect of source characteristics on task comparability in integrated writing tasks. *Assessing Writing*, 41, 25–46. <https://doi.org/10.1016/j.asw.2019.05.003>
- Huang, H.-T. D., & Hung, S.-T. A. (2018). Investigating the strategic behaviors in integrated speaking assessment. *System*, 78, 201–212. <https://doi.org/10.1016/j.system.2018.09.007>
- Huang, H.-T. D., Hung, S.-T. A., & Hong, H.-T. V. (2016). Test-taker characteristics and integrated speaking test performance: A path-analytic study. *Language Assessment Quarterly*, 13(4), 283–301. <https://doi.org/10.1080/15434303.2016.1236111>
- In'nami, Y., Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP®) test in relation to the TOEFL iBT® test. *Language Testing in Asia*, 6, Article 3. <https://doi.org/10.1186/s40468-016-0025-9>
- Ismayilli Karakoç, A., Gu, P., & Ruegg, R. (2025). Designing a rating scale for an integrated reading-writing test: A needs-oriented approach. *Assessing Writing*, 64, Article 100918. <https://doi.org/10.1016/j.asw.2025.100918>
- Jiang, L., Abbuhl, R., & Fu, Y. (2025). Effects of working memory and task type on syntactic complexity in EFL learners' writing. *International Review of Applied Linguistics in Language Teaching*, 63(4), 2695–2715. <https://doi.org/10.1515/iral-2023-0160>
- Jin, Z., & Webb, S. (2024). The effectiveness of note taking through exposure to L2 input: A meta-analysis. *Studies in Second Language Acquisition*, 46(2), 404–426. <https://doi.org/10.1017/S0272263123000529>
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133. <https://doi.org/10.1177/1558689806298224>
- Kim, A.-Y. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, 32(2), 227–258. <https://doi.org/10.1177/0265532214558457>

- Kormos, J., Brunfaut, T., & Michel, M. (2020). Motivational factors in computer-administered integrated skills tasks: A study of young learners. *Language Assessment Quarterly*, 17(1), 43–59. <https://doi.org/10.1080/15434303.2019.1664551>
- Kormos, J., Suzuki, S., & Eguchi, M. (2022). The role of input modality and vocabulary knowledge in alignment in reading-to-speaking tasks. *System*, 108, Article 102854. <https://doi.org/10.1016/j.system.2022.102854>
- Kunnan, A. J., Qin, C. Y., & Zhao, C. G. (2022). Developing a scenario-based English language assessment in an Asian university. *Language Assessment Quarterly*, 19(4), 368–393. <https://doi.org/10.1080/15434303.2022.2073886>
- Kyle, K. (2020). The relationship between features of source-text use and integrated writing quality. *Assessing Writing*, 45, Article 100467. <https://doi.org/10.1016/j.asw.2020.100467>
- Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019). Mapping master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, 41(3), 555–582. <https://doi.org/10.1017/S0272263119000251>
- Lestari, S. B., & Brunfaut, T. (2023). Operationalizing the reading-into-writing construct in analytic rating scales: Effects of different approaches on rating. *Language Testing*, 40(3), 684–722. <https://doi.org/10.1177/02655322231155561>
- Liao, X., Zhao, P., & Li, Z. (2025). The relationship between executive functions, source use, and integrated writing performance. *Assessing Writing*, 65, 100936. <https://doi.org/10.1016/j.asw.2025.100936>
- Liao, X., Zhu, X., & Cheong, C. M. (2021). Direct and indirect effects of independent language skills on the integrated writing performance of Chinese-speaking students with low proficiency. *Reading and Writing*, 34, 2529–2557. <https://doi.org/10.1007/s11145-021-10152-7>
- Lu, X., Révész, A., Michel, M., Kourtali, N., Lee, M., & Borges, L. (2025). Investigating the relationships of writing behaviours to linguistic complexity and accuracy in independent and integrated writing task performance. *Applied Linguistics Review*, 16(4), 1551–1579. <https://doi.org/10.1515/applirev-2024-0004>
- Machili, I., Papadopoulou, I., & Kantaridou, Z. (2020). Effect of strategy instruction on EFL students' video-mediated integrated writing performance. *Journal of Second Language Writing*, 48, 100708. <https://doi.org/10.1016/j.jslw.2019.100708>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Michel, M., Révész, A., Lu, X., Kourtali, N.-E., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research*, 36(3), 307–334. <https://doi.org/10.1177/0267658320915501>
- Min, S., & He, L. (2022). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, 39(1), 90–116. <https://doi.org/10.1177/0265532221995475>
- Mirzaei, A., Heidari Vincheh, M., & Hashemian, M. (2020). Retrofitting the IELTS reading section with a general cognitive diagnostic model in an Iranian EAP context. *Studies in Educational Evaluation*, 64, 100817. <https://doi.org/10.1016/j.stueduc.2019.100817>
- Nakamura, K. (2022). Investigating the factor structure of the Test of English for Academic Purposes (TEAP) and its relation to test takers' perceived test task value. *Language Testing in Asia*, 12, Article 35. <https://doi.org/10.1186/s40468-022-00183-4>

- Nelson, N., & King, J. R. (2023). Discourse synthesis: Textual transformations in writing from sources. *Reading and Writing*, 36(4), 769–808. <https://doi.org/10.1007/s11145-021-10243-5>
- Neumann, H., Leu, S., & McDonough, K. (2019). L2 writers' use of outside sources and the related challenges. *Journal of English for Academic Purposes*, 38, 106–120. <https://doi.org/10.1016/j.jeap.2019.02.002>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1–13. <https://doi.org/10.1177/1609406917733847>
- O'Grady, S., & Taşkesen, Ö. (2022). Developing a rating scale for integrated assessment of reading-into-writing skills. *Language Learning in Higher Education*, 12(1), 159–183. <https://doi.org/10.1515/cercles-2022-2049>
- Ohta, R., Plakans, L., & Gebril, A. (2018). Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. *Assessing Writing*, 38, 21–36. <https://doi.org/10.1016/j.asw.2018.08.001>
- Payant, C., McDonough, K., Uludag, P., & Lindberg, R. (2019). Predicting integrated writing task performance: Source comprehension, prewriting planning, and individual differences. *Journal of English for Academic Purposes*, 40, 87–97. <https://doi.org/10.1016/j.jeap.2019.06.001>
- Pu, P., & Chang, D. Y.-S. (2025). Effects of different input modes on blended EFL speaking instruction: A quasi-experimental study. *Computer Assisted Language Learning*, 38(5–6), 1223–1248. <https://doi.org/10.1080/09588221.2023.2273853>
- Ranalli, J., Feng, H.-H., & Chukharev-Hudilainen, E. (2018). Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing*, 36, 77–89. <https://doi.org/10.1016/j.asw.2018.03.007>
- Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, 41(3), 605–631. <https://doi.org/10.1017/S027226311900024X>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53. <https://doi.org/10.1177/0265532219871470>
- Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (2025). Individual differences in L2 listening proficiency revisited: Roles of form, meaning, and use aspects of phonological vocabulary knowledge. *Studies in Second Language Acquisition*, 47(1), 26–52. <https://doi.org/10.1017/S027226312300044X>
- Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. <https://doi.org/10.1177/1362168816643111>
- Sawaki, Y., & Sinharay, S. (2018). Do the TOEFL iBT section scores provide value-added information to stakeholders? *Language Testing*, 35(4), 529–556. <https://doi.org/10.1177/0265532217716731>
- Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2024). Developing and validating an online module for formative assessment of summary writing with automated content feedback for EFL academic writing instruction. *Language Testing in Asia*, 14, Article 50. <https://doi.org/10.1186/s40468-024-00325-w>
- Schoonen, R. (2019). Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. *Reading and Writing*, 32, 511–535. <https://doi.org/10.1007/s11145-018-9874-1>
- Shin, S.-Y., & Ewert, D. (2015). What accounts for integrated reading-to-write task scores? *Language Testing*, 32(2), 259–281. <https://doi.org/10.1177/0265532214560257>

- Siegel, J. (2018). Did you take “good” notes?: On methods for evaluating student notetaking performance. *Journal of English for Academic Purposes*, 35, 85–92. <https://doi.org/10.1016/j.jeap.2018.07.001>
- Siegel, J. (2020). Effects of notetaking instruction on intermediate and advanced L2 English learners: A quasi-experimental study. *Journal of English for Academic Purposes*, 46, 100868. <https://doi.org/10.1016/j.jeap.2020.100868>
- Siegel, J. (2024). From notes to writing: Three students in focus. *ELT Journal*, 78(3), 297–307. <https://doi.org/10.1093/elt/ccad022>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach’s alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Suzuki, S., & Kormos, J. (2023). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, 45(1), 38–64. <https://doi.org/10.1017/S0272263121000899>
- Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, 1(1), 77–100. <https://doi.org/10.1177/1558689806292430>
- Uchihara, T., & Clenton, J. (2023). The role of spoken vocabulary knowledge in second language speaking proficiency. *The Language Learning Journal*, 51(3), 376–393. <https://doi.org/10.1080/09571736.2022.2080856>
- Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, 47(1), 64–75. <https://doi.org/10.1080/09571736.2016.1191527>
- Uludag, P., Lindberg, R., McDonough, K., & Payant, C. (2019). Exploring L2 writers’ source-text use in an integrated writing assessment. *Journal of Second Language Writing*, 46, Article 100670. <https://doi.org/10.1016/j.jslw.2019.100670>
- Vandermeulen, N., van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2020). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing*, 33(2), 239–266. <https://doi.org/10.1007/s11145-019-09958-3>
- Vasylets, O., & Marín, J. (2021). The effects of working memory and L2 proficiency on L2 writing. *Journal of Second Language Writing*, 52, Article 100786. <https://doi.org/10.1016/j.jslw.2020.100786>
- Wang, J., Engelhard, G., Raczynski, K., Song, T., & Wolfe, E. W. (2017). Evaluating rater accuracy and perception for integrated writing assessments using a mixed-methods approach. *Assessing Writing*, 33, 36–47. <https://doi.org/10.1016/j.asw.2017.03.003>
- Wang, M., Gan, Q., & Boland, J. E. (2022). Effect of interactive intensity on lexical alignment and L2 writing quality. *System*, 108, Article 102847. <https://doi.org/10.1016/j.system.2022.102847>
- Wei, W., & Zheng, Y. (2017). An investigation of integrative and independent listening test tasks in a computerised academic English test. *Computer Assisted Language Learning*, 30(8), 864–883. <https://doi.org/10.1080/09588221.2017.1373131>
- Westbrook, C. (2023). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61, Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Wette, R. (2018). Source-based writing in a health sciences essay: Year 1 students’ perceptions, abilities and strategies. *Journal of English for Academic Purposes*, 36, 61–75. <https://doi.org/10.1016/j.jeap.2018.09.006>
- Xie, Q. (2023). Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing*, 57, 100755. <https://doi.org/10.1016/j.asw.2023.100755>

- Yan, X., Cheng, L., & Ginther, A. (2019). Factor analysis for fairness: Examining the impact of task type and examinee L1 background on scores of an ITA speaking test. *Language Testing*, 36(2), 207–234. <https://doi.org/10.1177/0265532218775764>
- Ye, W., & Ren, W. (2019). Source use in the story continuation writing task. *Assessing Writing*, 39, 39–49. <https://doi.org/10.1016/j.asw.2018.12.001>
- Ye, W., & Ren, W. (2023). Toward a better understanding of skill integration in integrated writing: A structural equation modeling study of EFL secondary learners' test performance. *Reading and Writing*, 36, 2739–2762. <https://doi.org/10.1007/s11145-023-10419-1>
- Yeager, R., Park, G., & Liao, R. J. T. (2024). Notetaking as validity evidence: A mixed-methods investigation of question preview in EAP listening assessment. *Journal of English for Academic Purposes*, 68, 101346. <https://doi.org/10.1016/j.jeap.2024.101346>
- Younas, A., Fàbregues, S., & Creswell, J. W. (2023). Generating metainferences in mixed methods research: A worked example in convergent mixed methods designs. *Methodological Innovations*, 16(3), 276–291. <https://doi.org/10.1177/20597991231188121>
- Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: A review of research (2011–2019). *Language Teaching*, 53(4), 491–523. <https://doi.org/10.1017/S0261444820000208>
- Zhao, P., Zhu, X., Yao, Y., & Liao, X. (2023). Ideal L2 self, enjoyment, and strategy use in L2 integrated writing: A self-regulatory learning perspective. *System*, 115, Article 103033. <https://doi.org/10.1016/j.system.2023.103033>
- Zhong, H. F. (2018). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*, 46(4), 357–370. <https://doi.org/10.1080/09571736.2015.1127403>
- Zhu, X., Li, G. Y., Cheong, C. M., Yu, G., & Liao, X. (2021). Secondary school students' discourse synthesis performance on Chinese (L1) and English (L2) integrated writing assessments. *Reading and Writing*, 34, 49–78. <https://doi.org/10.1007/s11145-020-10065-x>
- Zhu, X., Li, X., Yu, G., Cheong, C. M., & Liao, X. (2016). Exploring the relationships between independent listening and listening-reading-writing tasks in Chinese language testing: Toward a better understanding of the construct underlying integrated writing tasks. *Language Assessment Quarterly*, 13(3), 167–185. <https://doi.org/10.1080/15434303.2016.1210609>
- Zhu, X., Sun, Y., Liu, Y., Xu, W., & Cheong, C. M. (2025). Towards a better understanding of integrated writing performance: The influence of literacy strategy use and independent language skills. *Assessing Writing*, 64, 100922. <https://doi.org/10.1016/j.asw.2025.100922>