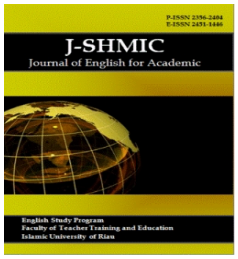

J-SHMIC : Journal of English for Academic

ISSN : 2641-1446 (Online) ISSN :2356-2404 (Print)
Homepage: https://journal.uir.ac.id/index.php/jshmic
Vol 13, No. 2, August 2026

A Needs Analysis of Field-Specific Listening Material Preferences among Vocational EFL Students

Muhammad Isa
English Study Program, Politeknik Negeri Madiun, Indonesia: muhammadisa@pnm.ac.id

ABSTRACT

Multimodal listening resources are often recommended for vocational English for Specific Purposes (ESP) classrooms, but studies of learner preference typically treat vocational learners as a single group, even though ESP theory holds that instruction should match the specific needs of each learner group rather than a generic one. This study compared three vocational tracks, hospitality, tourism, and content creation, on their general preference for multimodal input and on the specific listening content and challenges they report. Data came from a bilingual questionnaire comprising 10 Likert-scale items and 4 open-ended items, completed by 76 Critical Listening students in a D3 English Study Program in East Java, Indonesia. Track membership was inferred from students' open-ended responses, yielding a single-track subsample of 47 for group comparison (hospitality n=12, tourism n=16, content creation n=19). Kruskal-Wallis tests found no statistically significant differences between tracks on any preference item, indicating that a general preference for multimodal input was not detectably different across tracks in this sample. Content analysis of the open-ended responses, however, showed clearly different content preferences and reported challenges: hospitality students favored staff-guest service scripts, tourism students favored tour-guiding and travel-narration material, and content-creation students favored vlogs, interviews, and social media content, with content-creation respondents also reporting more connectivity and distraction-related difficulties. Together, these findings suggest a practical distinction for ESP course design: the format of listening instruction can remain multimodal across a shared course, while the content it carries may still need to be differentiated by occupational track.

KEYWORDS
ESP needs analysis; multimodal listening; vocational English; field-specific materials; listening challenges

ABSTRAK

Sumber belajar listening multimodal banyak direkomendasikan untuk kelas English for Specific Purposes (ESP) vokasi, tetapi penelitian mengenai preferensi pembelajar biasanya memperlakukan pembelajar vokasi sebagai satu kelompok, padahal teori ESP menekankan bahwa pengajaran seharusnya sesuai dengan kebutuhan spesifik setiap kelompok pembelajar, bukan kebutuhan yang generik. Penelitian ini membandingkan tiga bidang vokasi, yaitu perhotelan, pariwisata, dan content creation, dalam hal preferensi umum terhadap input multimodal serta konten dan tantangan listening spesifik yang mereka laporkan. Data diperoleh dari kuesioner dwibahasa, sepuluh item Likert dan empat item terbuka, yang diisi oleh 76 mahasiswa Critical Listening pada Program Studi D3 Bahasa Inggris di Jawa Timur, Indonesia. Keanggotaan bidang studi disimpulkan dari jawaban terbuka mahasiswa, menghasilkan subsampel bertrack tunggal sebanyak 47 mahasiswa untuk perbandingan kelompok (perhotelan n=12, pariwisata n=16, content creation n=19). Uji Kruskal-Wallis tidak menemukan perbedaan yang signifikan secara statistik antarbidang pada item preferensi mana pun, menunjukkan bahwa preferensi umum terhadap input multimodal tidak terdeteksi berbeda

KATA KUNCI
analisis kebutuhan ESP; listening multimodal; bahasa Inggris vokasi; materi spesifik bidang; tantangan listening

antarbidang pada sampel ini. Analisis konten terhadap jawaban terbuka, sebaliknya, menunjukkan preferensi konten dan tantangan yang jelas berbeda: mahasiswa perhotelan lebih menyukai skrip layanan staf-tamu, mahasiswa pariwisata lebih menyukai materi pemanduan wisata dan narasi perjalanan, dan mahasiswa content creation lebih menyukai vlog, wawancara, dan konten media sosial, dengan mahasiswa content creation juga melaporkan lebih banyak kesulitan terkait koneksi internet dan gangguan fokus. Secara keseluruhan, temuan ini menunjukkan adanya perbedaan praktis bagi desain kursus ESP: format pengajaran listening dapat tetap bersifat multimodal dalam satu kursus bersama, sementara konten yang dibawanya mungkin tetap perlu dibedakan menurut bidang pekerjaan.

***Corresponding Author:**
Muhammad Isa
muhammadisa@pnm.com
English Study Program
Politeknik Negeri Madiun, Indonesia

INTRODUCTION

Vocational English programs in Indonesia routinely enroll students preparing for distinct occupational fields within a single course. In the D3 English Study Program examined here, students in the Critical Listening course come from hospitality, tourism, and content-creation tracks, three fields with markedly different target communicative situations: hospitality centers on service encounters between staff and guests, tourism centers on guiding, narration, and traveler-facing exchanges, and content creation centers on publicly directed, often informal spoken media. Multimodal listening resources, materials that combine audio with images, video, or on-screen text, are widely recommended for such classrooms, but studies of learner preference typically treat the whole cohort as a single, undifferentiated group rather than asking whether preference and need differ by occupational field.

A companion needs-analysis study conducted with the same student population examined multimodal listening preference and motivation as a single, cohort-wide construct, drawing on self-determination theory and Mayer's (2021) cognitive theory of multimedia learning to explain why multimodal input sustains engagement. That companion study did not ask whether the three occupational tracks differ from one another, either in their general preference for multimodal input or in the specific listening content they consider relevant; it treated hospitality, tourism, and content-creation students as one group throughout. This gap matters because ESP theory does not merely recommend multimodal delivery in general, it also holds that content should answer each group's own target situation (Basturkmen, 2010; Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987), a claim the companion study's single-group design could not test. The present study addresses this gap directly. It reuses the same data-collection instrument as the companion study, a methodological link disclosed in full in the Data Source and Disclosure section below, but asks a different question of a different, larger extract of respondents: not whether multimodal input motivates listening in general, but whether the answer to that question, and the content that should accompany it, depends on

which occupational track a learner belongs to. The key contribution is therefore comparative: this is, to our knowledge, the first study at this institution to set hospitality, tourism, and content-creation students against one another within the same vocational EFL cohort, rather than examining each track, or the cohort as a whole, in isolation.

Three research questions guided the study. First, do students from the hospitality, tourism, and content-creation tracks differ in their general preference for multimodal listening resources? Second, what specific listening materials and content do students from each track report as most relevant to their perceived needs, and how does this preferred content differ across tracks? Third, do the listening challenges students report differ by track? Answering these questions extends the multimodal listening literature, which has mostly treated multimodal input as a generic instructional variable (Hsieh, 2020; Sydorenko, 2010; Wagner, 2010), by testing whether its value, and the content it should carry, is uniform or field-dependent within a single vocational cohort. It also extends prior ESP needs-analysis work in this institutional context, which has so far examined single tracks separately (Nastiti et al., 2026) or single vocational listening skills in isolation (Dwi Prastyo et al., 2023; Hadijah & Shalawati, 2018; Nurhasanah & Kurniawan, 2023; Rizqiani & Novitri, 2023), by comparing three tracks against one another within one dataset.

LITERATURE REVIEW

Multimodal Input and Listening Comprehension

Video-based input has repeatedly been shown to support comprehension differently from audio-only input (Wagner, 2010). Captioned video, similarly, supports vocabulary uptake alongside comprehension (Hsieh, 2020), and multimodal design more generally is grounded in a cognitive account of how learners process words and pictures together rather than in isolation (Mayer, 2021). Listening itself, independent of modality, is well documented as a demanding skill. Goh (2000) identified perception, parsing, and utilisation as three distinct stages at which comprehension can break down; Field (2008) catalogued the range of bottom-up and top-down problems learners face; and Vandergrift (2007) traced how the field has moved from testing listening toward teaching it explicitly. Taken together, this literature establishes multimodal input as a generally beneficial instructional variable, but it says little about whether that benefit, or the content it should carry, depends on who the learners are and what they are being prepared to do.

ESP and the Target Situation

Within English for Specific Purposes (ESP), the value of any instructional resource is judged not only by whether it supports general comprehension but by whether it answers the learner's target situation: the specific tasks, registers, and communicative demands that a given group of learners will face (Basturkmen, 2010; Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987; Hyland, 2022). A resource well matched to one occupational field's communicative demands is not automatically well matched to another. ESP needs analysis conventionally separates target needs, what learners need to be able to do in their eventual occupational setting, from learning needs, what helps them get there, and further separates necessities, lacks, and wants within target needs (Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987). On this framework, a preference for multimodal delivery sits on the learning-needs side of the distinction: it concerns how instruction is packaged so that learners can process it more easily, a question of pedagogic method rather than of the communicative

content a specific occupation demands. This classification is consistent with Mayer's (2021) claim that pairing words and pictures aids comprehension as a general property of how learners process information, not as a property tied to any particular occupational content; if the mechanism itself is content-independent, its instructional implication, that material should be delivered multimodally, should likewise hold regardless of field. The specific content of listening materials, guest-service dialogue versus tour narration versus vlog commentary, is by contrast a target-needs question: it concerns what the packaging should contain, and that answer is expected on ESP theory's own terms to vary by field. The RQ1 analysis below tests this classification empirically by asking whether the three tracks do, in fact, converge on multimodal preference while diverging on content.

Vocational English Needs Across Occupational Fields

Needs-analysis studies of Indonesian vocational English consistently confirm that different occupational fields carry distinct linguistic demands. Hospitality students and employees alike prioritize guest-facing service vocabulary and functions such as handling telephone enquiries and guest requests ((Kaharuddin et al., 2019), and tourism and hospitality students specifically name listening and speaking as their most needed skills for industry communication (Lertchalermtipakoon et al., 2021). Hospitality students elsewhere report unresolved communication difficulties even after completing dedicated coursework (Prabowo & Saptiany, 2024). A comparable pattern holds outside hospitality and tourism. Santika et al. (2022) examined a multimedia vocational programme rather than a content-creation program as such, but the two fields share a comparable profile: both center on producing publicly directed audiovisual material for an audience rather than on a face-to-face service exchange, which is why that study is treated here as informative for, though not direct evidence of, content-creation students' needs. In that study, students' most pressing needs centered on copywriting, technical vocabulary, and authentic materials for videography and photography rather than on generic English content (Santika et al., 2022), underscoring that occupational fields differ in content needs even when they share a vocational, technology-facing profile. Taken together, these studies establish that field-specific content needs exist, but each was conducted within a single occupational field; none has compared multimodal preferences, desired content, and listening challenges across multiple vocational fields drawn from the same cohort and instrument. Whether the field-specific differences documented separately in hospitality, tourism, and multimedia contexts also hold when hospitality, tourism, and content-creation students are asked the same questions side by side, and whether they extend beyond content to general pedagogical preferences such as multimodal delivery, is the gap this study addresses directly.

Read together, the three subsections above point to a single conceptual position. The literature on multimodal input reviewed first establishes multimodality as a broadly beneficial, seemingly content-independent instructional variable. ESP theory, reviewed next, explains why that content-independence is plausible for delivery format specifically, format is a learning-needs question, while insisting that content is a separate, target-needs question that should vary by field. The vocational-needs literature reviewed last then shows, field by field, that content demands do in fact diverge sharply by occupation, while offering no direct evidence about whether the same students, compared within one instrument, converge on delivery preference at the same time that they diverge on content. Research Questions 1 through 3 test this conceptual position directly: RQ1 asks whether the content-independent

mechanism proposed above actually holds across tracks in this cohort; RQ2 and RQ3 ask whether the field-specific divergence documented in the vocational-needs literature reappears, within the same cohort and instrument, as differences in desired content and reported challenges.

METHOD

Research Design

This study used a concurrent descriptive mixed-methods needs-analysis design, in which quantitative and qualitative strands were collected through a single instrument and analyzed separately before being integrated in the Discussion. The quantitative strand addressed RQ1 through between-track comparison of Likert-scale preference items; the qualitative strand addressed RQ2 and RQ3 through content analysis of open-ended responses. Collecting both strands concurrently, rather than sequentially, was intended to test whether process-level preference (RQ1) and content-level need (RQ2, RQ3) move together or separately across tracks within the same respondents at the same point in time.

Data Source and Disclosure

Data were collected through the same bilingual questionnaire used in a companion needs-analysis and motivation study conducted with the same student population; that companion study analyzes the data as a mixed-methods motivation study grounded in self-determination theory and multimedia learning theory, drawing on a subset of respondents together with follow-up interviews to explain why multimodal input motivates listening in general. The present study draws on a later, larger extract of the same ongoing data collection (N=76 at the time of this analysis), asks a distinct research question, whether preference and need differ by occupational track, and uses a different analytic approach, between-track group comparison and content analysis by track, rather than thematic analysis of interview data. Because the companion study neither disaggregates respondents by track nor reports the between-track comparisons presented here, the present analysis is not a re-reporting of the companion study's findings but a separate test of a question the companion study's single-group design did not address. This overlap in data source is disclosed here in the interest of transparency; no interview data and no motivation-theory analysis from the companion study are reproduced in this article.

Participants, Ethical Considerations, and Track Coding

Respondents were 76 students enrolled in the Critical Listening course of the D3 English Study Program at a state polytechnic in East Java, Indonesia (52 female, 24 male), responding to an anonymous, pseudonym-based Google Form.

Participation was voluntary, and respondents were informed of the study's purpose before completing the form. No personally identifying information was collected; responses were linked only to a self-selected pseudonym, and confidentiality was maintained throughout data storage and analysis.

The instrument did not include a dedicated background item recording each respondent's occupational track; track was therefore inferred from respondents' answers to the open-ended item asking what listening materials they would like to see more of, given their field of study (hospitality, tourism, or content creation). Responses that named exactly one track, or that described materials unambiguously specific to one track, were coded to that track (hospitality n=12, tourism n=16, content creation n=19; single-track subsample N=47). Responses naming

two or more tracks (n=19) and responses that did not identify a track (n=10) were retained in the full-sample descriptive statistics but excluded from the between-track comparison, since assigning them to a single group would misrepresent their content. Inferring track membership from an open-ended response in this way is a substantive limitation: it risks classification bias if respondents who mentioned a track did so incidentally rather than because that track defines their occupational focus, and it necessarily excludes over a third of the sample from the comparative analysis. A dedicated background item recording track membership directly, ideally verified against institutional enrollment records, would remove this source of bias in future data collection; this limitation is revisited in the Conclusion.

Instrument

The questionnaire comprised three parts. Part A recorded a pseudonym and sex. Part B consisted of ten statements rated on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree), covering exposure to subtitled and unsubtitled video, podcast use outside class, ease of comprehension with visual support, preference for multimodal over audio-only material, motivation from interactive tools, confidence with visual support, desire for more multimodal material in the course, technology's effect on practice time, and perceived match between current materials and future professional English use. Part C consisted of four open-ended items asking which resource type respondents found most motivating and why, a specific recalled experience with a multimodal or technology-assisted resource, the listening materials they would like to see more of considering their field of study, and the challenges they experience with technology-assisted listening resources. The instrument was presented bilingually in English and Bahasa Indonesia.

Data Analysis

Descriptive statistics (mean, standard deviation, and mean converted to a percentage of the five-point scale) were computed for the ten Likert items across the full sample (N=76). The percentage figure was calculated as the sample mean divided by the maximum possible scale score of 5, multiplied by 100; it is reported alongside the mean to give a more readily interpretable sense of how close responses were, on average, to maximum agreement, and should be read as a rescaling of the same mean rather than as an independent measure. For the single-track subsample (n=47), a Kruskal-Wallis H test was run on each of the ten items to test for between-track differences; this nonparametric test was chosen because Likert-scale data are ordinal and the three subgroups were small and unequal in size. Effect size for each Kruskal-Wallis comparison was calculated as epsilon-squared ($\varepsilon^2 = H / (N - 1)$, with N=47) and is reported alongside H and p in Table 2.

Open-ended responses to the items on desired materials and on challenges were content-analyzed using manifest-level categories. Categories were derived through a combination of deductive and inductive coding: broad categories (for example, service content, guiding content, and media content) were anticipated from the occupational literature reviewed earlier, while their specific labels (for example, guest check-in and check-out, tour-guide commentary, vlogs and interviews) were refined inductively from the wording respondents actually used. Desired materials were grouped by the specific content named and tabulated by track; challenges were coded into six non-mutually-exclusive categories (speaker speed or accent, audio or visual clarity, internet connectivity, vocabulary gaps, distraction or focus, and none reported), with a response eligible for more than one code where it described more than one difficulty, and cross-tabulated by track. Coding was performed by a single researcher; a second

coder was not available to establish inter-coder reliability, which is treated as a limitation of the qualitative strand and is discussed further in the Conclusion.

FINDING AND DISCUSSION

FINDING

Overall Multimodal Preference Profile

Across the full sample (N=76), agreement with the ten multimodal preference items ranged from 65.3% to 88.4% of the maximum possible score (Table 1). The strongest agreement was with the general preference for multimodal over audio-only material (Q5, M=4.42, 88.4%), followed by the desire for more multimodal material in the course itself (Q8, M=4.12, 82.4%) and reported confidence when visual support is present (Q7, M=4.08, 81.6%). The weakest agreement, though still above the scale midpoint, concerned independent practice: watching English videos without subtitles (Q2, M=3.26, 65.3%) and using podcasts outside class (Q3, M=3.30, 66.1%) both trailed the general stated preference for multimodal material. This gap between stated preference and independent practice recurs across the two related studies conducted with this population and is consistent with the broader finding in the literature that ease of comprehension and habitual independent use, while related, are distinct outcomes (Hsieh, 2020).

Table 1. Descriptive Statistics for Multimodal Preference Items, Full Sample (N=76)

Q1 Watch subtitled video	3.79 (0.96)	75.8	Q6 Interactive tools motivate	4.04 (0.96)	80.8
Q2 Watch unsubtitled video	3.26 (1.12)	65.3	Q7 Confident with visual support	4.08 (0.93)	81.6
Q3 Use podcasts outside class	3.30 (1.08)	66.1	Q8 Want more multimodal in CL	4.12 (0.99)	82.4
Q4 Easier with visuals present	4.01 (0.95)	80.3	Q9 Technology increases practice time	3.99 (0.79)	79.7
Q5 Prefer multimodal over audio-only	4.42 (0.96)	88.4	Q10 Materials match future profession	3.83 (0.93)	76.6

Track Comparison of Multimodal Preference (RQ1)

Within the single-track subsample (n=47), mean scores on the ten Likert items were broadly similar across the hospitality, tourism, and content-creation groups (Table 2). Kruskal-Wallis tests found no statistically significant difference between tracks on any item (all $p>.10$), including the two items most directly tied to the course itself: the desire for more multimodal material in Critical Listening (Q8: $H=1.08$, $p=.582$) and perceived match between current materials and future professional use (Q10: $H=0.54$, $p=.762$).

Table 2. Track Comparison of Multimodal Preference Items, Single-Track Subsample (n=47)

Item	Hospitality (n=12)	Tourism (n=16)	Content Creation (n=19)	H	p	Effect size (ε^2)
Q1	3.75	3.94	3.58	1.56	.458	.034
Q2	2.83	3.50	3.37	2.87	.238	.062
Q3	3.25	3.62	2.89	3.76	.153	.082
Q4	4.17	3.94	4.00	0.16	.921	.003
Q5	4.50	4.69	4.11	4.16	.125	.090

Q6	4.42	3.75	4.11	3.39	.184	.074
Q7	3.83	4.12	3.89	0.30	.861	.007
Q8	3.83	4.31	3.84	1.08	.582	.023
Q9	4.08	4.06	3.79	1.18	.553	.026
Q10	3.83	3.69	3.84	0.54	.762	.012

Note. Effect size is epsilon-squared, a rank-based effect size for the Kruskal-Wallis test calculated as $\varepsilon^2 = H / (N - 1)$, where $N = 47$. Values are small across all ten items (ε^2 ranging from .003 to .090), reinforcing that the absence of statistical significance reflects a genuinely small between-track difference in this sample rather than a large effect obscured by low statistical power.

This null result should be interpreted cautiously: with subgroups this small and unequal, a non-significant difference indicates that no difference was detected in this sample, not that the three tracks' preferences are equivalent in the wider population (see the Participants, Ethical Considerations, and Track Coding section and the Conclusion for a fuller discussion of this limitation). The accompanying effect sizes (Table 2, ε^2 = .003 to .090) are uniformly small, which at least suggests the non-significant result reflects a genuinely small between-track difference in this sample rather than a large effect that the small subgroups were simply underpowered to detect. With that caveat, the pattern is at least consistent with Mayer's (2021) account of multimodal processing as a general, content-independent mechanism, and with intervention evidence that pairing authentic video with interactive quizzes raised vocational EFL listening scores regardless of specialization (Negara & Jamilah, 2025). In ESP terms, this pattern is consistent with treating the multimodal-versus-audio-only preference as a learning-needs question rather than a target-needs one (Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987), the classification proposed in the ESP and the Target Situation section above: on the present evidence, how these learners learn best does not appear to depend on which occupational field they are preparing for, though a larger and more balanced sample would be needed to state this with confidence. The tentative practical implication, developed further in the Synthesis section below, is that a single Critical Listening course may be able to deliver its content multimodally for all three tracks without varying the format by track.

Field-Specific Content Needs (RQ2)

Where the quantitative preference items showed convergence, the qualitative responses on desired materials showed clear divergence by track (Table 3). Hospitality-track respondents most often named service-encounter content: staff-guest conversations, check-in and check-out exchanges, complaint handling, and receptionist dialogue, echoing one respondent's request for material 'such as conversations between receptionist and guest by phone, in person, and handling problems.' Tourism-track respondents instead named guiding and travel-narration content: tour-guide commentary, price negotiation, and traveler-facing exchanges. Content-creation-track respondents converged on publicly directed, informal media: vlogs, podcast interviews, brand promotional videos, and social media clips, with one respondent explaining the wish to 'analyze real-world materials like brand promotional videos, lifestyle vlogs, podcast interviews, and trendy social media clips, for practical insight into communication and audience engagement.'

Table 3. Track-Specific Desired Listening Materials

Hospitality (n=12)	Staff-guest service scripts: guest check-in and check-out, complaint handling, telephone and in-person exchanges with receptionists, and customer service dialogue.
Tourism (n=16)	Tour-guiding and travel narration: tour-guide commentary, price negotiation, travel and destination talk, plain conversational exchanges with tourists.
Content Creation (n=19)	Vlogs, interviews, and social media content: brand promotional videos, lifestyle vlogs, podcast interviews, content-creator commentary, trendy social media clips.

These patterns are broadly consistent with each field's documented communicative demands, though the studies cited below were conducted with different samples and should be read as corroborating rather than confirming the present findings. Hospitality's request for service scripts parallels the priority hospitality-track students at the same institution place on guest-interaction and service communication in a separate needs analysis (Nastiti et al., 2026), aligns with employees' own ranking of listening and speaking as the most important English skills for hotel work (Prima, 2022), and echoes the specific functions, using the telephone, receiving guests, handling requests, that other Indonesian hospitality needs analyses identify as core content (Kaharuddin et al., 2019). Both this study and the broader literature point toward the same practical conclusion: vocational hospitality instruction should be built around scenario-based, guest-facing dialogue rather than general-purpose material, and communication difficulties can persist even after coursework where that content match is missing (Prabowo & Saptiany, 2024). Tourism's request for guiding and narration content reflects the fact that a tour guide's spoken output is typically extended, one-directional commentary rather than the two-party service exchange that defines hospitality encounters, a distinction also visible in survey data showing tourism and hospitality students both naming listening among their most-needed industry skills even as their preferred content diverges (Lertchalermtipakoon et al., 2021); tourism materials that model narration and destination description therefore appear better matched to this track's needs than service-dialogue material would be. Content creation's request for vlogs, interviews, and social media clips reflects a target situation defined not by a workplace service encounter but by public-facing, often unscripted spoken media, a profile that parallels, without directly replicating, findings from a needs analysis of a multimedia vocational programme, where students' priority content was copywriting and authentic video and photography material rather than generic English (Santika et al., 2022; see the discussion of the comparability of this field to content creation above). This parallel may help explain why content creation's stated preferences diverge furthest from the other two tracks. Read together with the tentative null result reported above for RQ1, this divergence supports treating the process of multimodal delivery and the content that delivery carries as two separable design decisions: the present data suggest the former can be standardized across the shared course, while the latter cannot.

Track-Specific Challenge Profiles (RQ3)

Reported listening challenges also varied by track, though less uniformly than desired content (Table 4). Vocabulary gaps were reported by 42% of hospitality respondents and 32% of content-creation respondents but by none of the tourism respondents in the subsample.

Internet connectivity problems were reported most by content-creation respondents (37%); distraction or loss of focus while listening was reported exclusively within this track (16%), and no other track reported it at all. Audio and visual clarity problems were reported most by hospitality respondents (42%). Speaker speed and accent were the most frequently reported challenge overall and were reported most by content-creation respondents (42%). Given the small subgroup sizes, these differences are reported here as descriptive tendencies within this sample rather than as tested, generalizable differences.

Table 4. Track-Specific Challenge Profiles, Single-Track Subsample (n=47)

Challenge category	Hospitality (n=12)	Tourism (n=16)	Content Creation (n=19)
Speaker speed / accent	4 (33%)	2 (13%)	8 (42%)
Audio / visual clarity	5 (42%)	5 (31%)	3 (16%)
Internet connectivity	3 (25%)	4 (25%)	7 (37%)
Vocabulary gaps	5 (42%)	0 (0%)	6 (32%)
Distraction / focus	0 (0%)	0 (0%)	3 (16%)
None reported	0 (0%)	3 (19%)	1 (5%)

Some tentative interpretation is possible where it connects directly to what respondents described. Content creation's elevated connectivity and distraction complaints are plausible given that this track's preferred materials, social media clips and streamed interviews, are themselves consumed through the same data-dependent platforms respondents named as distracting; this is also consistent with broader survey evidence that heavier personal device and internet use among students correlates with measurable differences in language-literacy outcomes (Saripi et al., 2024), though the present study did not measure device use directly and this specific connection remains speculative for this sample. Hospitality's clarity concerns are consistent with this track's emphasis on short, exchange-based dialogue, where a single unclear utterance (a room number, a request) can derail comprehension of the whole exchange, though this too is offered as a plausible reading of the pattern rather than a tested claim. Speaker speed and accent, the single most frequently reported challenge overall, is well documented outside this sample as a persistent source of difficulty even for otherwise proficient L2 listeners (Field, 2008; Goh, 2000), and unfamiliar accents specifically have been shown to lower both self-reported comprehensibility and measured intelligibility relative to more familiar speech (Verbeke & Simon, 2023); this is consistent with, though not direct evidence for, content creation's especially high rate of this challenge, which we tentatively attribute to that track's stated preference for unscripted, native-paced social media speech. Beyond this, we do not draw further conclusions about the specific causes of the vocabulary-gap or clarity patterns, since the present instrument did not ask respondents to explain why they experienced each challenge.

Synthesis: Separating Process from Content in ESP Listening Design

Taken together, the findings from RQ1 through RQ3 point to a single, practically useful distinction, framed here as a tentative pattern that future work with larger samples should confirm. The process question, whether listening material should be delivered multimodally rather than as audio alone, produced no significant differences across the three tracks in this sample and can provisionally be answered once for the whole Critical Listening cohort: a

shared multimodal instructional format appears defensible for all three tracks. This is compatible with the format choices already used in the same institutional and journal context, where YouTube-based delivery improved vocational speaking outcomes regardless of specialization (Ilyas & Putri, 2020) and podcast-based listening instruction was positively received across a mixed cohort (Dwi Prastyo et al., 2023; Hadijah & Shalawati, 2018). The content question, what that multimodal material should be about, produced clearly track-specific answers that a shared format cannot resolve on its own. A single set of subtitled videos and interactive quizzes, however well designed, is unlikely to meet a hospitality student's need for service-encounter dialogue, a tourism student's need for guiding narration, and a content-creation student's need for vlog and interview material at the same time, even though all three students might respond equally well to that material being multimodal rather than audio-only; nor is content the only field-specific dimension, since even the technical vocabulary a track needs to master has been shown to vary enough across ESP populations to warrant separate treatment (Herdiansyah & Hardiyanto, 2026). Concretely, this could mean retaining a single subtitled-video-plus-quiz format across the Critical Listening course while rotating the video content by week or by track, for example, using a hotel check-in scenario, a guided-tour recording, and a short vlog segment as parallel materials for the same listening sub-skill, so that all three groups practice the same skill through content matched to their own target situation. This is precisely the distinction that ESP theory draws between learning needs and target needs (Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987); the present data suggest, tentatively, that this distinction is not only theoretical but potentially visible within a single vocational cohort responding to the same instrument, a possibility that warrants confirmation with a larger and more balanced sample.

CONCLUSION

This study compared hospitality, tourism, and content-creation track students in a vocational Critical Listening course (N=76) on their general preference for multimodal listening resources and on the specific listening content and challenges they report. Quantitatively, the three tracks did not differ significantly on any of the ten multimodal preference items in this sample; given the small and unequal subgroup sizes discussed below, this absence of a detected difference should not be read as evidence that the tracks' preferences are equivalent, only that no difference was found here. Qualitatively, the three tracks named clearly different desired materials, service-encounter dialogue for hospitality, guiding and travel narration for tourism, and vlogs, interviews, and social media content for content creation, and showed differing, track-linked challenge profiles. The central contribution of this study is the distinction these two results draw jointly: instructional format (multimodal, technology-assisted delivery) may reasonably be held constant across a shared vocational course, while the content units within that format still need to be differentiated by occupational track, for example, by rotating hospitality, tourism, and content-creation material across weeks within the same subtitled-video-and-quiz format described in the Synthesis section above.

This study has several limitations, the most consequential of which concern how track membership was determined and how the qualitative data were coded. Track membership was inferred from an open-ended response rather than measured with a dedicated survey item, which required excluding respondents who named more than one track or none and introduces a risk of classification bias in the remaining subsample. The single-track subgroups used for

between-track comparison were, in consequence, small and unequal in size, ranging from 12 respondents in the hospitality subgroup to 19 in the content-creation subgroup, which limits the statistical power of the between-track comparisons. The study was also cross-sectional, confined to one course at one institution, and its open-ended responses were coded by a single researcher without independent verification. Future research should build a dedicated track-membership item into the instrument from the outset, ideally verified against institutional records; recruit larger and more balanced track subgroups, possibly by pooling across cohorts or institutions; and test the curriculum implication proposed in the Synthesis section directly, for example through a comparative or experimental study contrasting learning outcomes when listening content is track-matched versus generic within an otherwise identical multimodal format. Findings should, in the meantime, be read as specific to this vocational cohort and institutional context rather than generalized to vocational EFL learners more broadly.

REFERENCES

Basturkmen, H. (2010). Developing courses in English for specific purposes. Palgrave Macmillan.

Dudley-Evans, T., & St John, M. J. (1998). Developments in English for specific purposes: A multi-disciplinary approach. Cambridge University Press.

Dwi Prastyo, Y., Dianingsih, A., & Farhana, S. (2023). Students' perception of using podcasts to improve listening skills at the 3rd semester students of English Department at Universitas Bandar Lampung. J-SHMIC: Journal of English for Academic, 10(2), 175–181. https://doi.org/10.25299/jshmic.2023.vol10(2).11177

Field, J. (2008). Listening in the language classroom. Cambridge University Press.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. System, 28(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Hadijah, S., & Shalawati, S. (2018). Enhancing English language learners' listening comprehension through instruction in listening strategies. J-SHMIC: Journal of English for Academic, 5(1). https://doi.org/10.25299/jshmic.2018.vol5(1).1204

Herdiansyah, Y. K., & Hardiyanto, A. (2026). Writing with precision: How Filipino ESL learners navigate technical terms and context clues. J-SHMIC: Journal of English for Academic, 13(1), 27–39. https://doi.org/10.25299/jshmic.2026.vol13(1).26656

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. Computer Assisted Language Learning, 33(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Hutchinson, T., & Waters, A. (1987). English for Specific Purposes: A learning-centred approach. Cambridge University Press.

Hyland, K. (2022). English for specific purposes: What is it and where is it taking us? ESP Today, 10(2), 202–220. https://doi.org/10.18485/esptoday.2022.10.2.1

Ilyas, M., & Putri, M. E. (2020). YouTube channel: An alternative social media to enhance EFL students' speaking skill. J-SHMIC: Journal of English for Academic, 7(1), 77–87. https://doi.org/10.25299/jshmic.2020.vol7(1).4141

Kaharuddin, Hikmawati, & Arafah, B. (2019). Needs analysis on English for vocational purpose for students of hospitality department. KnE Social Sciences, 3(19), 344–387. https://doi.org/10.18502/kss.v3i19.4869

Lertchalermtipakoon, P., Wongsubun, U., & Kawinkoonlasate, P. (2021). Need analysis: English language use by students in the tourism and hospitality industry. English Language Teaching, 14(3), 59–71. https://doi.org/10.5539/elt.v14n3p59

Mayer, R. E. (2021). Multimedia learning (3rd ed.). Cambridge University Press.

Nastiti, I. A., Noviabahari, J. L., & Prakosha, D. (2026). Initial need analysis of applied English for students of English study program in the hospitality field. PAEDAGOGIA, 29(1), 67–77. https://doi.org/10.20961/paedagogia.v29i1.114774

Negara, M. H. J., & Jamilah. (2025). Improving English listening skills for vocational students through interactive ICT tools. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(2), 5360–5377. https://doi.org/10.24256/ideas.v13i2.8133

Nurhasanah, T., & Kurniawan, E. (2023). ESP need analysis of computer and network engineering in vocational high school. J-SHMIC: Journal of English for Academic, 10(2), 139–154. https://doi.org/10.25299/jshmic.2023.vol10(2).13347

Prabowo, B. A., & Saptiany, S. G. (2024). Communication challenges in the hospitality business: Analysis of students' strategies in learning English at STIEPARI Semarang. LITE: Jurnal Bahasa, Sastra, Dan Budaya, 20(1), 74–85. https://doi.org/10.33633/lite.v20i1.10216

Prima, S. (2022). A study of perception of the importance of English language skills among Indonesian hotel employees. J-SHMIC: Journal of English for Academic, 9(1), 73–86. https://doi.org/10.25299/jshmic.2022.vol9(1).8972

Rizqiani, D. A., & Novitri, S. (2023). Developing needs analysis of critical literacy models for teaching EFL reading class. J-SHMIC: Journal of English for Academic, 10(2), 94–107. https://doi.org/10.25299/jshmic.2023.vol10(2).12831

Santika, S., Wirza, Y., & Emilia, E. (2022). A needs analysis for ESP in a multimedia study programme of a vocational high school. International Journal of Education, 15(1), 61–68. https://doi.org/10.17509/ije.v15i1.46156

Saripi, S., Asari, & S. (2024). Evaluating the effectiveness of PBL in enhancing speaking skills in non-formal education. https://doi.org/10.25134/erjee.v12i3.11166

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. Language Learning & Technology, 14(2), 50–73.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. Language Teaching, 40(3), 191–210. https://doi.org/10.1017/S0261444807004338

Verbeke, G., & Simon, E. (2023). Listening to accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech. Lingua, 292, 103572. https://doi.org/10.1016/j.lingua.2023.103572

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. Language Testing, 27(4), 493–513. https://doi.org/10.1177/0265532209355668