

COMPARATIVE ANALYSIS OF GRAMMATICAL DIFFICULTIES IN TOEFL SWE SECTION ACROSS DISCIPLINARY BACKGROUNDS: A MIXED-METHODS STUDY OF INDONESIAN EFL LEARNERS

¹Ari Arifin Danuwijaya, ²Abu Nawas

¹Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No.229, Isola, Bandung City, West Java, Indonesia

²School of Education, The University of Adelaide, Adelaide Australia Selatan 5005, Australia

*Corresponding Author Email: aridanuwijaya@upi.edu

Article Info	Abstract
Article History Received: May 2025 Revised: June 2025 Accepted: September 2025 Published: October 2025	<i>Grammatical competence remains fundamental to English proficiency assessment, particularly in standardized tests such as the TOEFL ITP. The Structure and Written Expressions (SWE) section consistently challenges EFL learners, yet research has predominantly focused on English language majors, with limited comparative examination of how academic disciplinary background influences grammatical performance and learner perceptions. This mixed-methods study addressed this gap by investigating grammatical difficulties across three academic programs: Tourism Education, Japanese Language Education, and English Language Education. The study examined 163 Indonesian undergraduate students' performance on the TOEFL ITP SWE section and explored their self-perceived competence and test-related challenges. Quantitative data from a 40-item TOEFL SWE test were analyzed using Classical Test Theory and jMetrik software to calculate item difficulty indices. Qualitative data were collected through self-rating questionnaires and focus group discussions with 15 participants. Results revealed systematic performance differences across disciplines, with Tourism and Japanese Language students demonstrating significantly higher error rates than English majors. Six grammatical constructions consistently emerged as problematic: relative adverbs, causal expressions, inverted sentence structures, word forms, parallel structures, and redundancy. Qualitative analysis identified shared challenges including vocabulary limitations, test format unfamiliarity, and time management difficulties. Notably, English students, despite higher performance, showed unexpected weaknesses in error analysis. The findings demonstrate that grammatical performance in standardized contexts results from complex interactions between linguistic competence, test literacy, and academic background.</i>
Keywords TOEFL ITP; Structure and written expressions; Grammatical competence; Item difficulty analysis;	
How to cite: Danuwijaya, A.A., & Nawas, A. (2025). Comparative Analysis of Grammatical Difficulties in TOEFL SWE Section Across Disciplinary Backgrounds: A Mixed-Methods Study of Indonesian EFL Learners. <i>JOLLT Journal of Languages and Language Teaching</i> , 13(4), 2040-2054. Doi: https://doi.org/10.33394/joltt.v13i4.15550	

INTRODUCTION

In the context of English as a Foreign Language (EFL) education, grammatical competence remains a core indicator of language proficiency, especially in high-stakes assessments such as the TOEFL ITP test. As one of the components of TOEFL ITP, the Structure and Written Expressions (SWE) section poses persistent challenges for learners. The section tests grammatical knowledge and the ability to identify subtle structural errors in decontextualized sentences under time pressure. Despite its diagnostic value, this section often receives less instructional emphasis in EFL classrooms than in reading or listening components (Rahmah, 2021; Ananda, 2016). As the TOEFL test continues to be widely used for academic

placement, scholarship eligibility, and graduation requirements in Indonesia, it has become a high-stakes assessment tool influencing educational and career opportunities (Djojonegoro, 2020; Rahma et al., 2022; Widiati & Cahyono, 2006). For many Indonesian students, achieving a minimum TOEFL score is more than just a benchmark. It is a gateway to university programs, overseas study opportunities, and employment. Given the high stakes, especially in the EFL context, understanding learners' specific challenges in the grammar section becomes crucial. With this understanding, educators can craft more focused and meaningful teaching strategies that genuinely meet students where they are and help them build the confidence and skills they need to succeed.

Research into learners' performance in TOEFL grammar sections has highlighted several recurring grammatical problems. Prior studies have identified frequent errors in subject-verb agreement, parallel structure, verb form, redundancy, and relative clauses (Akmal et al., 2020; Mahmud, 2014). Tilana and Yunita (2019) found that redundancy and misplaced modifiers were particularly problematic for students, even among those majoring in English. In a similar vein, Hajri et al. (2018) found that students often struggled the most with recognizing errors in clause structures and conjunctions. This shows that many EFL students still feel unprepared when faced with the specific demands of the TOEFL, and they often lack the strategies and test familiarity needed to tackle this section with confidence and accuracy.

While the challenges faced by English majors in grammar-based assessments have been extensively discussed, there is still limited research comparing the grammatical performance of students across non-English academic disciplines, such as Tourism or Japanese Language Education. Studies that include students from various academic programs can offer more nuanced insights into how academic background influences language acquisition and grammatical competence. Zuhayana (2018) observed that non-English majors struggle more with sentence structure and lexical choice, mainly due to minimal exposure to English grammar in their core curriculum. A study by Hasan (2019) supports this approach, showing that interference from the first language (L1), negative learning experiences, and limited exposure to English are the main factors contributing to structural errors in the TOEFL.

However, what remains critically under-examined is how grammar instruction differs across academic disciplines. While English Education programs typically incorporate explicit grammar teaching with metalinguistic awareness and test preparation strategies, Tourism Education curricula tend to prioritize communicative competence for hospitality contexts with minimal formal grammar instruction. Japanese Language Education programs present a unique case, where students develop sophisticated metalinguistic awareness through learning Japanese grammar but may lack equivalent analytical frameworks for English structures. This variation in grammar pedagogy creates different preparation for standardized assessments, yet few studies have systematically contrasted these instructional approaches and their outcomes. Furthermore, much of the previous research has focused solely on test scores and error types without incorporating learners' perceptions, self-evaluations, and test-taking experiences, which are critical for understanding the cognitive and affective dimensions of language assessment (Yosintha et al., 2021; Nurhayati & Nehe, 2016).

This study addresses these gaps by examining and comparing the grammatical difficulties encountered by students from three different academic programs in completing the SWE section of the TOEFL ITP. Using a mixed-methods approach, the study combines item-level quantitative analysis with qualitative data from student self-assessments and focus group discussions. By combining what the test scores show with how students feel and think about the questions, this approach offers a more complete picture of the challenges they face. It also looks at which specific grammar points students find hardest, using item difficulty analysis, a valuable yet often overlooked tool in language testing research in Indonesia. This is helpful to understand better where different groups of learners tend to struggle the most.

This study aimed to examine how academic disciplinary background influences grammatical performance and learner perceptions in TOEFL contexts through key objectives: identifying the most frequent grammatical errors and challenging items across different academic programs, determining which TOEFL grammar constructions demonstrate the highest difficulty levels empirically, exploring how students from various disciplines evaluate their grammatical competence and test-taking experiences, and developing evidence-based pedagogical recommendations tailored to specific disciplinary needs. Specifically, this research seeks to answer the following questions: 1) What are the most frequent grammatical errors made and difficult items shown by students from Tourism Education, Japanese Language Education, and English Language Education when completing the TOEFL ITP SWE section? 2) How do students from different academic programs perceive their grammatical competence and test-taking challenges in the TOEFL SWE section? and 3) What pedagogical implications emerge from the interaction between grammatical performance, test literacy, and academic discipline in TOEFL preparation? Through analysis of performance patterns and learner perceptions, this research contributes to improving grammar instruction and TOEFL preparation strategies in Indonesian higher education settings, with broader implications for EFL contexts serving diverse academic populations.

Grammatical Competence in EFL and Standardized Testing

Grammatical competence is the mastery of syntactic rules and the ability to apply them in real-time communication, which is central to communicative competence (Canale & Swain, 1980). In EFL contexts, this competence underpins learners' success in productive skills and high-stakes standardized tests such as the TOEFL ITP. The Structure and Written Expressions (SWE) component presents unique challenges by isolating decontextualized sentences and requiring test-takers to identify single underlined errors, a task that presupposes both declarative grammatical knowledge and procedural analytical skills (Brown, 2004). However, many EFL instructional programs emphasize discrete point grammar teaching, which is often removed from authentic usage and leaves students ill-equipped for the cognitive demands of error detection under time pressure (Rahmah, 2021; Nurhayati & Nehe, 2016).

This assessment approach reveals a critical pedagogical tension in EFL instruction. Traditional grammar teaching in many EFL programs emphasizes discrete-point instruction focused on rule memorization rather than analytical sentence processing, leaving students inadequately prepared for the cognitive demands of error detection under time constraints (Rahmah, 2021; Nurhayati & Nehe, 2016). Research consistently demonstrates that students struggle most with complex grammatical constructions precisely because they lack systematic practice in parsing and reassembling sentences (Putra, 2020; Tanihardjo, 2017). Putra's (2020) findings on word order and parallelism difficulties, combined with Tanihardjo's (2017) observations about insufficient training in identifying inverted structures, emphasize the need for instructional approaches that mirror the analytical demands of standardized assessments.

This pedagogical gap suggests that effective TOEFL preparation requires moving beyond traditional grammar instruction toward cognitively demanding tasks such as sentence combining and systematic error analysis (Corder, 1967). The implications extend beyond test preparation to fundamental questions about how grammatical competence should be developed and assessed in EFL contexts.

Patterns of Difficulty in TOEFL SWE

Empirical investigations across multiple Indonesian EFL contexts reveal consistent patterns of grammatical difficulty that transcend individual institutional differences. Subject-verb agreement, verb forms, connectors and conjunctions, reduced clauses, inversions, and parallel structures emerge repeatedly as the most problematic categories (Ananda, 2016; Syahrin, 2020; Bulan et al., 2023; Ridwan et al., 2024). These patterns suggest systematic gaps

in grammatical instruction rather than isolated learning difficulties. Akmal et al.'s (2020) mixed-methods investigation provides particularly valuable insights into the distribution of errors across SWE components. Their findings that learners most frequently struggle with determiners, conjunction placement, and reduced clauses in Structure items, while facing difficulties with adverb connectors and concession clauses in Written Expression, reveal section-specific challenges that require targeted instructional responses. The consistency of these patterns across different studies indicates that certain grammatical constructions pose universal challenges for Indonesian EFL learners, regardless of institutional context.

Particularly concerning is the frequent overlooking of redundancy and parallel structure in grammar curricula, despite their prominent appearance in TOEFL assessments (Tilana & Yunita, 2019). Slamet and Sulistyaningsih's (2021) finding that 55% of TOEFL-like test items proved challenging, with vocabulary limitations compounding grammatical weaknesses, highlights the multifaceted nature of SWE difficulties. This suggests that successful performance requires not only grammatical knowledge but also sufficient vocabulary breadth and test-taking strategies. The persistent mismatch between classroom instruction and assessment demands reflects broader tensions in EFL pedagogy. The TOEFL SWE's emphasis on discrete grammatical knowledge over communicative authenticity has been criticized for potentially encouraging form-focused rather than meaning-focused learning (Zhuang, 2008; Raimes, 1990). However, the practical reality that students must navigate these assessments for academic and professional advancement necessitates instruction that addresses both communicative competence and test performance requirements (Etika, 2020; Handayani, 2019).

Methodological Approaches to Analyzing Test Performance

Item difficulty analysis provides crucial diagnostic information that remains underutilized in classroom-based TOEFL research. As a fundamental psychometric measure representing the proportion of test-takers answering an item correctly, item difficulty indices reveal which constructions most effectively differentiate learner abilities (Brown, 2005; Meyer, 2014). The application of Classical Test Theory principles through software like jMetrik enables systematic identification of problematic grammatical areas that might otherwise escape pedagogical attention.

Research consistently demonstrates that items with difficulty indices below 0.30 correspond to complex structures inadequately represented in standard grammar curricula (Hampp et al., 2021; Ridwan et al., 2024). This empirical evidence provides objective foundation for curricular decisions about instructional emphasis and time allocation. The standard interpretation framework, which is items below 0.30 as "difficult," 0.30-0.70 as "moderate," and above 0.70 as "easy", offers practical guidance for educators seeking to prioritize instructional focus based on empirical evidence rather than intuition (Brown, 2005).

Beyond pedagogical applications, item difficulty analysis can identify potential assessment design issues, including ambiguous wording or excessive lexical complexity that may confound grammatical measurement (Ebel & Frisbie, 1991). This dual diagnostic function that informs both instruction and assessment development positions item difficulty analysis as an essential tool for improving EFL grammar education.

Disciplinary Background and Language Assessment Performance

The influence of academic disciplinary background on grammatical performance represents a significantly understudied area with substantial implications for EFL instruction. Most existing TOEFL SWE research focuses exclusively on English language majors, creating a knowledge gap about how students from other academic disciplines experience grammatical assessment (Nurhayati & Nehe, 2016; Zuhayana, 2018). This limitation is particularly

concerning given the increasing numbers of non-English majors required to demonstrate English proficiency for academic and professional purposes.

Non-English majors, including those in Tourism and Japanese Language Education, typically receive limited grammar instruction beyond general English courses, resulting in systematically lower confidence levels and higher error rates in SWE tasks (Ratnaningsih, 2022; Ramadhika & Riadil, 2022). Zuhayana's (2018) investigation of mathematics and management students revealed particular struggles with idiomatic expressions and complex clause structures, directly attributable to minimal academic English integration in disciplinary curricula. These findings suggest that grammatical difficulties may be as much a function of curricular exposure as inherent learning challenges.

Conversely, English majors, despite generally stronger performance, face distinct challenges with nuanced grammatical distinctions such as restrictive versus non-restrictive relative clauses and comparative/superlative error identification (Mahmud, 2014; Thiel, 2019). This pattern indicates that even extensive grammatical instruction may inadequately prepare students for the analytical precision required in standardized assessments.

The implications extend beyond individual performance to broader questions of educational equity and access. If grammatical assessment performance varies systematically by disciplinary background due to differential curricular exposure rather than inherent ability differences, then standardized language assessments may inadvertently disadvantage students from certain academic fields. This possibility highlights the importance of developing differentiated instructional approaches that acknowledge diverse disciplinary backgrounds while maintaining appropriate academic standards. The present study addresses this critical gap by systematically comparing grammatical performance and perceptions across three distinct academic programs, providing empirical foundation for evidence-based instructional differentiation in EFL contexts.

RESEARCH METHOD

Research Design

This study used a mixed-methods design that combines a quantitative descriptive approach with an exploratory qualitative component. The main objective of this study was to identify and compare the types of grammatical errors found in the SWE section of the TOEFL ITP test among students from three academic programs: Tourism Education, Japanese Language Education, and English Language Education. The quantitative approach was used to classify and quantify grammatical errors and to analyse item difficulty. In contrast, the qualitative approach explored students' perceptions of the problems they experienced when completing this section of the TOEFL test.

Participants

The participants of this study were 163 undergraduate students enrolled in the first year of their respective programs at a university in Bandung, Indonesia. The sample comprised 55 students from Tourism Education, 53 from Japanese Language Education, and 55 from English Education. All participants had completed at least one semester of English topics. Before the data were collected, informed consent was obtained from all participants. The participants were informed about the purpose and procedures of the study and their voluntary participation. They were assured that their involvement was confidential, their responses would be anonymous, and they could withdraw from the study at any point without consequence.

Instruments

The primary instrument used in this study was a complete set of questions from the SWE section of the TOEFL ITP, taken from The Official Guide to the TOEFL ITP Test (ETS, 2013). The test consisted of 40 multiple-choice items, comprising 15 from Part A (Structure)

and 25 from Part B (Written Expression). This widely used test was selected because it has been internationally recognized for its accuracy and reliability. In this study, its internal consistency was measured using Cronbach's Alpha, resulting in a score of 0.76, which shows an acceptable level that shows the test performs reliably in language learning contexts (Brown, 2004). A wide range of student performance from a lowest score of 6 to a highest of 35, with an average score of 17.69 was identified in the study. This variation highlighted how differently students from each academic program approached the test, and thus, the exploration and comparison of their performance more closely is needed.

Data Collection Procedure and Analysis

In this study, data were collected in two systematic phases. In the first phase, participants completed the SWE section of the TOEFL ITP under timed conditions (25 minutes), simulating a standardized test environment. The test was administered as a diagnostic instrument during the initial session of their English course. Upon completion, students were instructed to transfer their answers into a Google Form, which enabled efficient digital data compilation and minimized transcription errors during subsequent analysis.

The analytical framework of this study was grounded in Classical Test Theory (CTT), which provided the basis for evaluating the psychometric properties of the TOEFL SWE test items. To analyze the test results more effectively, this study used jMetrik, a free and widely used software developed by Meyer (2014) for educational research. jMetrik makes examining how test items perform easier by offering helpful tools for analyzing questions, checking reliability, reporting scores, and visualizing data clearly and meaningfully. In this study, it was used to calculate item difficulty indices, defined as the proportion of correct responses per item, and to estimate the internal consistency reliability of the test using Cronbach's Alpha. Item difficulty values ranged from 0 (most difficult) to 1 (easiest) and were interpreted using classical benchmarks recommended by Brown (2005): items with values below 0.30 were considered difficult, 0.30–0.70 moderate, and above 0.70 easy. The analysis focused on item difficulty as an indicator of how challenging specific grammatical structures were for students across the three academic programs.

In addition to the CTT-based statistical analysis, the study included a post-hoc linguistic classification of incorrect responses. Error categories, such as subject-verb agreement, verb tense, connectors and conjunctions, word order, and pronoun usage, were based on established grammar taxonomies (e.g., Pyle, 2001). This complemented the item analysis by providing insights into the nature of the grammatical challenges encountered by students. This study also provides a complete understanding of both how the test items work and the linguistic challenges students face.

In the next phase, qualitative data were gathered to add depth and context to the quantitative results, helping to provide a complete picture of the learners' experiences. First, all 163 participants provided a self-assessment score (0–10) reflecting their perceived knowledge of English grammar, alongside open-ended reflections on their difficulties with the TOEFL test. Additionally, 15 students (5 from each program) were selected based on convenience sampling to participate in Focus Group Discussions (FGDs). These FGDs explored students' perceptions of the most challenging question types, their learning experiences in grammar, and their strategies during the test.

Qualitative data were analyzed using thematic analysis, beginning with transcription and coding, then identifying key themes such as lack of vocabulary, unfamiliarity with test format, time pressure, and reliance on intuition. The combination of methodological approaches yields a more comprehensive understanding of the linguistic errors and cognitive challenges students from diverse academic backgrounds face. In addition to that, it also highlights how student experience and academic discipline intersect in shaping grammatical proficiency in TOEFL contexts.

FINDINGS AND DISCUSSION

RQ1: Most Frequent Grammatical Errors and Difficult Items Across Academic Programs

This section directly addresses the first research question by presenting quantitative analysis of error patterns and item difficulty indices across the three academic programs. The analysis provides a picture of grammatical challenges across different academic disciplines and examines students' perceptions and struggles. The results of this study revealed notable patterns of difficulty encountered by students from three academic disciplines, i.e., Tourism Education, Japanese Language Education, and English Language Education, in responding to the SWE section of the TOEFL ITP test. Across the 163 participants, the error rate remained considerably high, with substantial variation observed across the groups. A closer inspection of item difficulty indices provided more insights into which grammatical areas posed the most persistent challenges and how these varied by program of study. Table 1 compares item difficulty indices in SWE across academic programs.

Table 1
Comparison of Item Difficulty Indices in SWE Across Academic Programs

Question	Topic	Tourism Education	Japanese Education	English Education	Notes
<i>Structure Part</i>					
5	Relative adverb	0.10	0.05	0.10	Use of where, when, why
6	Causal expressions	0.12	0.16	0.03	Use of <i>because</i> or <i>because of</i>
13	Inverted sentence structure	0.12	0.17	0.17	Use of Subject-Verb order
<i>Written Expressions Part</i>					
27	Word form	0.05	0.15	0.15	Adjective vs Noun Usage
37	Parallel structure	0.18	0.18	0.25	Parallelism in word formation
38	Redundancy	0.16	0.15	0.27	Use of conjunction

Note. Item difficulty is based on classical test theory, which calculates the proportion of correct responses per item. Values below 0.30 indicate difficult items (Brown, 2005)

In the Structure section, the analysis identified three items as consistently difficult across all three groups, directly answering the research question about the most frequent problematic areas. Item 5 (relative adverb), item 6 (causal expressions), and item 13 (inverted sentence structure) emerged as the most challenging constructions. Item 5 was identified as the most challenging, with item difficulty indices of 0.10 (Tourism), 0.05 (Japanese), and 0.10 (English), indicating that fewer than 10% of students, on average, were able to answer correctly. This suggests a widespread lack of familiarity with relative adverbs such as where, when, and why, possibly due to limited exposure to these structures in academic discourse. Similarly, item 6 on causal expressions (e.g., because vs because of) demonstrated the lowest index of 0.03 among English Education students, highlighting the surprising fact that even students with formal training in English grammar struggle to identify logical relations in complex sentences.

The systematic error analysis reveals distinct disciplinary patterns that directly address the research question about program-specific difficulties. Tourism Education students showed the most pronounced difficulties with lexical discrimination tasks, particularly Item 27 (word form) with a difficulty index of 0.05, reflecting their limited exposure to academic English vocabulary and morphological awareness. Their errors frequently involved basic verb forms, connectors, and word class distinctions, indicating gaps in foundational grammatical knowledge. Japanese Language Education students exhibited consistent moderate performance across items but struggled particularly with relative constructions and tense-related structures,

likely due to syntactic interference from their L1 and different grammatical frameworks between Japanese and English. Most unexpectedly, English Education students demonstrated surprising difficulty with causal expressions (Item 6: 0.03 difficulty index), suggesting that even students with formal linguistic training struggle with nuanced logical connectors and advanced grammatical distinctions.

In the Written Expressions section, cross-group analysis highlighted a common vulnerability in recognizing lexical and structural errors, particularly in items involving word form, redundancy, and parallel structure. For example, item 27, which tested word form accuracy, recorded difficulty indices of 0.05 (Tourism), 0.15 (Japanese), and 0.15 (English), indicating that even English Education students had difficulty distinguishing between parts of speech in context. Similarly, items 37 (parallel structure) and 38 (redundancy) were persistently difficult across groups, with average difficulty indices below 0.30.

These findings echo previous research patterns while providing new insights into disciplinary variations. The persistent errors observed across these diverse academic groups echo the broader trend documented in prior research, which highlights that structural complexity, syntactic unfamiliarity, and limited exposure to test-specific formats play critical roles in EFL learners' underperformance (Slamet & Sulistyaningsih, 2021; Halim & Ardiningtyas, 2018). One notable point of convergence between this study and previous research is the difficulty with inversion, connectors, relative clauses, and verb forms, which appear across the error profiles of all three student groups. These areas were similarly identified in Putra (2020), who found word choice, verb form, and parallelism to be recurring sources of error. Akmal et al. (2020) also emphasized students' difficulty with determiners, reduced clauses, and conjunctions, particularly when dealing with complex syntactic constructions, which aligns closely with the problematic items identified in the present study, such as relative adverbs and causal expressions.

Furthermore, the issue of redundancy, which appeared prominently in the current research findings, especially in Written Expressions, is mirrored in Tilana and Yunita's (2019) study, which identified redundancy as the most difficult grammatical element in TOEFL Structure for English department students. This study's multi-disciplinary approach reveals that error patterns are not universal but systematically related to students' academic backgrounds and curricular exposure, corroborating Zuhrayana's (2018) findings about the impact of discipline-specific language preparation on grammatical performance.

RQ2: Student Perceptions of Grammatical Competence and Test-Taking Challenges

This section directly addresses the second research question through comprehensive analysis of self-rating data and qualitative reflections, revealing significant differences in how students from different academic programs perceive their grammatical competence and experience test-related challenges. To complement the quantitative results, this study collected qualitative data through students' self-assessments and open-ended reflections regarding their experiences with the TOEFL SWE section. The participants were asked to rate their knowledge on a scale of 0 (no knowledge at all) to 10 (perfect mastery) and describe any difficulties they encountered while answering the test. Table 2 summarizes the self-rating data and key qualitative themes that emerged from student reflections across three academic programs.

Table 2
Summary of Student Self-Ratings and Qualitative Themes by Academic Programs

Academic Program	Self-Rating Range	Mean Self-Rating	Primary Themes	Quotes
Tourism Education	0-3	1.08	- Vocabulary limitations - Complete unfamiliarity with test format - Lack of preparation	"I didn't understand the questions at all."

Academic Program	Self-Rating Range	Mean Self-Rating	Primary Themes	Quotes
			- Overwhelming confusion	"I lack vocabulary." "I have never practiced questions like these."
Japanese Language Education	2-5	3.2	- Test format unfamiliarity - Complex sentence confusion - Limited confidence - Grammar rule uncertainty	"I didn't understand the questions" "The sentences were too complex" "I'm not confident with English grammar."
English Language Education	4-9	6.1	- Confusion with similar options - Parallel structure challenges - Redundancy identification issues - Strategic analysis gaps	"All the underlined parts look correct." "I had difficulty with parallel structure." "I answered based on feeling."

Note. Self-ratings were collected on a 0-10 scale where 0 = no knowledge and 10 = perfect mastery. Qualitative themes were identified through thematic analysis of open-ended reflections and focus group discussions

The quantitative self-assessment data showed distinct patterns that directly answer the research question about perceived competence across programs. The self-rating data revealed a clear pattern consistent with the students' academic backgrounds. Students from the Japanese Language Education program generally rated themselves between 2 and 5, showing limited confidence in their understanding of English grammar. Many reported unfamiliarity with test-specific grammar formats and admitted to being confused by complex sentence structures. Tourism Education students rated themselves even lower on average, with scores commonly falling between 0 and 2. Their comments showed a lack of preparedness, with expressions such as "I didn't understand the questions at all," "I lack vocabulary," and "I have never practiced questions like these." On the other hand, English Education students provided more moderate ratings, typically between 5 and 7, with a few scoring 8 or 9. Despite their relatively higher self-assessments, several still admitted to confusion when answering similar-looking options or identifying subtle grammatical errors, particularly in parallel structure and redundancy.

Students' reflections revealed several recurring themes across the three groups, providing detailed insights into the qualitative differences in their test-taking experiences. Many non-English majors reported that unfamiliar vocabulary significantly hindered their understanding. They expressed feelings such as, "I don't know the meaning of the words," "The vocabulary is difficult," and "I can't understand the sentences." These gaps in vocabulary often hindered their ability to analyse grammatical structures effectively. The second theme identified was a lack of familiarity with the test format. Many students expressed that this was their first experience with such questions, with comments like, "I've never taken a grammar test like this before," and "I don't know how to answer TOEFL-style questions."

Error analysis challenges emerged as a particularly notable theme for English students (45.5%), directly addressing the research question about program-specific challenges. They frequently commented on the difficulty of identifying errors when "all options seem correct." Many noted frustrations like:

"Menentukan kata yang error karena ada beberapa soal yang kelihatannya tidak ada error jadi harus dianalisis lebih dalam lagi" (Determining which word is wrong because some questions seem to have no errors so they need to be analyzed more deeply) (Student 116, English language student)

“Error analyzing menurut saya cukup membingungkan, ada beberapa kesalahan yang cukup terlihat, ada beberapa yang tidak” (Error analyzing is quite confusing to me, some errors are quite visible? some are not). (Student 137, English language student)

Time management represented another significant challenge that varied systematically by program, affecting 27.3% of English students, 21.8% of Tourism students, but only 5.7% of Japanese students. Students across programs expressed feeling rushed during the test, with comments ranging from *“Waktu nya”* (The time), *“Time”*, to more detailed explanations like *“I panicked and couldn’t manage my time well so there’s some questions I answered randomly.”*

“mungkin waktu terlalu singkat jadi tak ada waktu untuk berfikir dan jadi terburu buru” (maybe the time was too short so there was no time to think and I became rushed). (Student 142, English language student)

Test unfamiliarity appeared as a concerning theme across all programs (10-15% of students), but its impact varied significantly by discipline. Many expressed that this was their first encounter with TOEFL-style questions. Particularly telling were comments like:

“Lumayan sulit karna pertama kali mengerjakan soal seperti ini” (Quite difficult because this is the first time doing questions like this) (Student 25, Japanese language student)

“karna tidak terbiasa mengerjakan soal seperti itu jadi pusing pak” (because I’m not used to doing questions like that, it makes me dizzy, sir). (Student 5, Japanese language student)

The analysis reveals that perceived competence does not always correlate with actual performance patterns, particularly in advanced error detection tasks. Interestingly, even students from the English Education program, who rated themselves moderately higher (ranging from 5 to 7 on average), reported significant challenges with error analysis, parallel structure, and redundancy. Many noted confusions when *“all the underlined parts look correct,”* which explains why these items remained difficult even among advanced learners. This finding aligns with Mahmud (2014), who observed that English majors, while generally better in grammar, still require strategic awareness and exposure to authentic test formats.

These findings demonstrate that test-taking challenges result from the complex interaction of linguistic competence, test literacy, and disciplinary background. This observation aligns with the high error rates in items involving word form and prepositional phrase usage. Students often described being overwhelmed by vocabulary, which affected comprehension and their ability to identify grammatical errors, especially in the Written Expressions section. This supports findings from Zhuang (2008) and Yosintha et al. (2021) who emphasized vocabulary as a mediating factor in structural comprehension. A recurrent issue mentioned across all groups was test anxiety under time constraints. This indicates that processing load and test pressure significantly influenced performance. These constraints may have compounded the difficulties students faced when dealing with more syntactically complex sentence structures, corroborating the cognitive load hypothesis as discussed in Halim and Ardiningtyas (2018).

RQ 3: Pedagogical Implications from the Interaction of Performance, Test Literacy, and Academic Discipline

The integration of quantitative data with qualitative self-assessment provides a comprehensive foundation for addressing the third research question about pedagogical implications. The findings from the qualitative data suggest that students’ performance relies not only on their grammatical competence but also on additional factors such as lexical access, cognitive strategies, exposure, and test literacy. These results highlight the necessity for tailored instructional design that ensures structural accuracy, enhances vocabulary depth, and implements effective testing strategies, all customized to the foundational aspects of each academic discipline.

The clear disciplinary patterns documented in this study directly inform evidence-based pedagogical recommendations that acknowledge the distinct learning needs emerging from different academic backgrounds. The analysis suggests that the grammatical difficulties encountered in the TOEFL are not limited to any discipline but vary according to academic background. The findings highlight the need for differentiated grammar instruction that addresses the general needs of EFL learners, and the syntactic challenges encountered by students from different academic fields.

For Tourism Education programs, the predominance of vocabulary limitations and basic grammatical confusion indicates the need for foundational language development. Tourism students demonstrated difficulty with verb tenses and connectors and showed the highest difficulty in verb-related structures and conjunctions, which corroborates Zuhayana's (2018) findings that non-English majors often face more pronounced challenges due to a lack of curriculum alignment with academic English demands. These students would benefit from scaffolded approaches that build from basic sentence structures toward more complex grammatical constructions, with explicit attention to word formation and lexical choice. Explicit vocabulary instruction through morphological awareness training has been shown to significantly improve both receptive and productive language skills (Nation, 2001; Schmitt, 2000), particularly for students with limited academic English exposure.

Japanese Language Education students' specific challenges with English syntactic features suggest the need for contrastive analysis approaches that explicitly address differences between Japanese and English grammatical systems. Japanese Language students struggled with relative clauses, reduced clauses, and infinitive constructions, a trend that may stem from structural differences between Japanese and English syntax. Their metalinguistic awareness could be leveraged through comparative grammar instruction that builds on their understanding of Japanese structures while highlighting English-specific features like relative clauses, tense systems, and logical connectors.

English Language Education students' difficulties with error analysis and strategic application of grammatical knowledge indicate the need for advanced test preparation that emphasizes analytical skills and strategic awareness. English Education students made fewer errors overall, though they faced challenges with more complex grammatical structures, such as parallelism, determiner usage, and redundancy. This indicates that while formal grammar instruction benefits English majors, there is still a need for targeted attention to advanced grammar skills and stylistic refinement. Despite their formal training, these students require explicit instruction in error detection techniques, parallel structure analysis, and redundancy identification, combined with timed practice that develops automatic processing of complex grammatical relationships. Research by Phakiti (2003) and Zhang (2001) demonstrates that explicit strategy instruction significantly improves performance on grammar-focused language tests, particularly for advanced learners who possess substantial declarative knowledge but lack strategic application skills.

The variation in test literacy across programs reveals a fundamental curricular gap that requires systematic reform. The persistent unfamiliarity with test formats across all programs highlights inadequate integration of assessment literacy within language instruction. This reflects the pedagogical gap identified by Nurhayati and Nehe (2016), where traditional grammar instruction fails to translate into standardized test performance. Curriculum reform should integrate authentic test practice throughout language instruction, moving beyond discrete grammar drills toward contextualized application that mirrors assessment demands (Ananda, 2016; Zhuang, 2008). This approach is supported by Bachman and Palmer's (1996) framework of language test usefulness, which emphasizes the importance of authenticity and interactivity in language assessment preparation. Research by Green (2007) and Saville and Hawkey (2004) on test preparation effects demonstrates that systematic exposure to test formats

and strategies can significantly improve performance without compromising the validity of test score interpretations.

The persistent theme of unfamiliarity with test formats across all programs highlights a fundamental pedagogical gap. Many expressed that they had never encountered those questions or had only practiced general grammar rules without contextual application. This reflects the pedagogical gap highlighted by Nurhayati and Nehe (2016), that traditional grammar instruction often does not translate well into test performance, particularly in standardized formats like TOEFL. This finding supports the need for curriculum reform that integrates authentic test practice throughout language instruction, moving beyond discrete grammar drills toward contextualized application that mirrors assessment demands.

These findings suggest that the SWE section functions not merely as a grammar test but as an assessment of strategic test awareness, linguistic adaptability, and disciplinary literacy. The systematic patterns identified through jMetrik-supported analytics, as distinguished from previous qualitative studies like Yosintha et al. (2021), provide empirical evidence for differentiated instructional design. Effective TOEFL preparation demands understanding how linguistic competence, test literacy, and academic background interact to shape performance, enabling educators to design targeted interventions that address the specific challenges emerging from different disciplinary contexts. This approach moves beyond one-size-fits-all grammar instruction toward evidence-based pedagogical practices that acknowledge the complex relationship between disciplinary preparation and language assessment performance, supporting recent calls for more contextualized and learner-centered approaches to language testing and preparation (McNamara & Roever, 2006; Chalhoub-Deville, 2003).

CONCLUSION

This study explored and compared the grammatical difficulties encountered by students from three academic programs when completing the SWE section of the TOEFL ITP test. The quantitative findings indicated a high overall error rate, particularly among students from non-English majors, with Tourism Education students recording the highest percentage of incorrect responses. Specific grammatical features, such as relative adverbs, causal expressions, verb forms, inversion structures, and word forms, emerged as the most problematic across all groups. Item difficulty analysis confirmed that several test items fell into the “difficult” category, with some items answered correctly by fewer than 10% of participants.

Qualitative data from self-assessment and focus group discussions provided further depth to these findings. Students reported low confidence in their grammar knowledge, unfamiliarity with test formats, lexical limitations, and cognitive overload under time constraints. Even English Education students, despite higher test scores and self-ratings, expressed challenges with nuanced grammar aspects such as redundancy, parallel structure, and determiner usage. These results emphasize that grammar difficulties in TOEFL are shaped by linguistic competence and test literacy, exposure, and academic discipline.

This research adds to understanding how academic background, test familiarity, and perceived competence affect grammatical performance in standardized testing contexts. The study provides diagnostic insight that can inform assessment practices and curriculum design by identifying error patterns and uncovering underlying cognitive and experiential factors. It also bridges a gap in the literature by examining non-English majors, which is an understudied group in TOEFL preparation research, thereby broadening the relevance of grammar-focused investigations in English for Academic Purposes (EAP).

The findings of this study have several implications for language teaching, assessment design, and institutional policy. First, TOEFL preparation programs should be tailored according to students’ disciplinary backgrounds, prioritizing grammatical areas that are empirically shown to be most difficult. Second, educators should integrate authentic test-like grammar practice early in instruction, especially for students in non-language fields. Third,

grammar teaching should move beyond rule memorization and emphasize contextual analysis, test strategies, and metacognitive skills. At the institutional level, language support services should consider diagnostic testing as a tool to identify and support students at risk of underperforming in academic English tasks.

REFERENCES

- Akmal, S., Rasyid, M. N. A., Masna, Y., & Soraya, C. N. (2020). EFL learners' difficulties in the structure and written expression section of TOEFL test in an Indonesian university. *Englisia: Journal of Language, Education, and Humanities*, 7(2), 156–180. <https://doi.org/10.23887/jpbi.v9i1.33811>
- Ananda, R. (2016). Problems with section two ITP TOEFL test. *Studies in English Language and Education*, 3(1), 35-49. <https://doi.org/10.24815/siele.v3i1.3387>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Brown, J. D. (2005). *Testing in language programs: A Comprehensive guide to English language assessment*. McGraw-Hill.
- Bulan, A., Wahyuni, N., Suryaningsih, L., & Hasan, H. (2023). Exploring EFL students' difficulties and strategies in learning TOEFL ITP of structure and written expression. *VELES (Voices of English Language Education Society)*, 7(1), 66–74. <https://doi.org/10.29408/veles.v7i1.6881>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369-383. <https://doi.org/10.1191/0265532203lt264oa>
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching*, 5(1–4), 161–170.
- Djojonegoro, W. (2020). Standardized English proficiency tests and their implications in Indonesian higher education. *Journal of English Language Pedagogy and Practice*, 13(1), 45–59.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
- Etika, I. (2020). Investigating the difficulties faced by English Department students in structure and written expression section of TOEFL at Bung Hatta University. *Lingua Litera*, 5(1), 42–69. <https://doi.org/10.55345/stba1.v5i1.46>
- ETS (2013). *Official guide to the TOEFL ITP test*. Educational Testing Service
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Hajri, T., Jufrizal, & Wahyuni, D. (2018). An analysis of difficulties in answering structure and written expression of TOEFL made by English students of Universitas Negeri Padang. *Journal of English Language Teaching*, 7(1). <https://doi.org/10.24036/jelt.v7i1.8957>
- Halim, N., & Ardiningtyas, S. Y. (2018). Difficulties faced by the students in answering TOEFL test questions. *ETERNAL: English, Teaching, Learning, and Research Journal*, 4(2), 219–231. <https://doi.org/10.24252/Eternal.V42.2018.A7>
- Hampp, P. L., Kumayas, T. A., & Lengkoan, F. (2021). Synthesizing grammar and structure problems faced by Indonesian TOEFL participants. *Jurnal Pendidikan Bahasa Inggris Undiksha*, 9(1), 64–68. <https://doi.org/10.23887/jpbi.v9i1.33811>

- Handayani, F. (2019). Structure and written expression section on paper-based TOEFL: Perceived difficulties by nursing students of Poltekkes Solok, West Sumatera. *Jurnal Educative: Journal of Educational Studies*, 4(2), 135–149.
- Hasan, B. (2019). Interlanguage in error analysis study. *Aksara: Jurnal Bahasa dan Sastra*, 30(2), 201–210.
- Mahmud, M. (2014). The EFL students' problems in answering the Test of English as a Foreign Language (TOEFL): A study in Indonesian context. *Theory & Practice in Language Studies*, 4(12). <https://doi.org/10.4304/tpls.4.12.2581-2587>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.
- Meyer, J. P. (2014). *jMetrik (Version 4.1.1) [Computer software]*. Retrieved from <https://www.jmetrik.com>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nurhayati, N., & Nehe, B. (2016). An analysis of students' strategies in answering TOEFL. *Journal of English Language Studies*, 1(1), 53–65. <https://doi.org/10.30870/jels.v1i1.682>
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26–56. <https://doi.org/10.1191/0265532203lt243oa>
- Putra, A. (2020). An analysis on English students' difficulties in TOEFL test of structure and written expression section at IAIN Padangsidimpuan (Unpublished undergraduate thesis). IAIN Padangsidimpuan.
- Pyle, M. A. (2001). TOEFL CBT. IDG Books Worldwide, Inc.
- Rahma, E. A., Syafitri, R., Oktavinanda, G., Rahmatillah, R., & Syahputri, V. N. (2022). Tertiary students' strategies and difficulties in answering the TOEFL test. *SALTel Journal (Southeast Asia Language Teaching and Learning)*, 5(2), 33–37. <https://doi.org/10.35307/saltel.v5i2.94>
- Rahmah, M. (2021). Teaching grammar in TOEFL preparation classes: A neglected skill? *Indonesian EFL Journal*, 7(1), 1–10. <https://doi.org/10.25134/ieflj.v7i1.3962>
- Raimes, A. (1990). The TOEFL test of written English: Causes for concern. *TESOL Quarterly*, 24(3), 427–442. <https://doi.org/10.2307/3587228>
- Ramadhika, B., & Riadil, I. G. (2022). Constraints on structure and written expressions of the TOEFL for midwifery students. *Metathesis: Journal of English Language, Literature, and Teaching*, 6(2), 199–212. <https://doi.org/10.31002/metathesis.v6i2.208>
- Ratnaningsih, P. W. (2022). Students' perceptions of structure and written expressions comprehension in TOEFL. *NIVEDANA: Jurnal Komunikasi dan Bahasa*, 3(2), 164–170. <https://doi.org/10.53565/nivedana.v3i2.671>
- Ridwan, R., Dewi, R. S., & Nappu, S. (2024). An analysis of graduate students' challenges in answering TOEFL structure and writing questions. *International Journal of Social Science and Human Research*, 07(12). <https://doi.org/10.47191/ijsshr/v7-i12-64>
- Sari, M. A. (2022). English students' difficulties with parallel structure in the structure and written expression section in TOEFL test. *Journal of English Language Teaching*, 11(1), 1–10.
- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73–96). Lawrence Erlbaum Associates.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Slamet, J., & Sulistyaningsih, S. (2021). Students' difficulties in answering "Structure and Written Expression" TOEFL-like at STKIP PGRI Sidoarjo. *E-Structural*, 4(01), 17–27. <https://doi.org/10.33633/es.v4i01.4410>

- Syahrin, A. (2020). *Error analysis in completing structure section of TOEFL test* (Undergraduate thesis, UIN Ar-Raniry). UIN Ar-Raniry Repository. <https://repository.ar-raniry.ac.id/id/eprint/13307/>
- Tanihardjo, J. (2017). The analysis of students' English competence in the grammar section in the paper-based TOEFL: A case study at English Department in Bunda Mulia University. *Journal of English Language and Culture*, 6(1). <http://dx.doi.org/10.30813/jelc.v6i1.270>
- Thiel, J. E. V. (2019). *An analysis on the students' problems in the structure and written expression section of TOEFL* (Undergraduate thesis). UNIKA Soegijapranata Semarang.
- Tilana, P. E., & Yunita, W. (2019). Students' English structure and written competence. *Journal of English Education and Teaching*, 3(2), 180–194. <https://doi.org/10.33369/jeet.3.2.180-194>
- Widiati, U., & Cahyono, B. Y. (2006). The teaching of EFL listening in the Indonesian context: The state of the art. *TEFLIN Journal*, 17(2), 169–179.
- Yosintha, R., Yuniarti, S. S., & Ramadhika, B. (2021). Structure and written expressions of the TOEFL: Linguistic and non-linguistic constraints. *NOBEL: Journal of Literature and Language Teaching*, 12(1), 70–90. <https://doi.org/10.15642/NOBEL.2021.12.1.70-90>
- Zhang, L. J. (2001). Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an acquisition-poor environment. *Language Awareness*, 10(4), 268–288. <https://doi.org/10.1080/09658410108667039>
- Zhuang, X. (2008). Practice on assessing grammar and vocabulary: The case of the TOEFL. *Online Submission*, 5(7), 46–57.
- Zuhayana, Z. (2018). Students' difficulties in answering structure and written expression questions of TOEFL. *Jurnal Serambi Ilmu*, 19(1), 1–11. <https://doi.org/10.32672/si.v19i1.989>