

Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation

Agung Pranoto Kadiatmaja

Pendidikan Bahasa Arab, STAI Ali bin abi Thalib, Surabaya, Indonesia



Email Korespondensi: agungpranotokadiatmaja@stai-ali.ac.id

INFO ARTIKEL

Histori Artikel:

Diterima 25-09-2026
Disetujui 29-09-2026
Diterbitkan 01-10-2026

ABSTRACT

This study investigates whether diagnostic-driven syntactic training, as opposed to conventional drilling, functions as a metacognitive scaffolding mechanism for 20 EFL learners preparing for the TOEFL ITP Structure and Written Expression subtest. Situated within a multilingual higher education context where English certification serves as a critical gateway for academic mobility, the program adopted a mixed-methods design across four sequential phases: diagnostic baseline testing, instructional strategy workshop, syntactic error analysis and remedial drilling, and post-test evaluation. Data were collected from pre- and post-tests, error mapping sheets, observation logs, and feedback surveys, and analyzed using the Shapiro-Wilk test, paired-sample t-test, N-gain, Dulay et al.'s (1982) Surface Strategy Taxonomy, and Braun and Clarke's (2006) thematic analysis. Findings indicate that diagnostic-driven training significantly improved participants' scores ($t(19) = 8.42, p < 0.001$, Cohen's $d = 1.89$), with a mean gain of 15.8 points and a medium N-gain of 0.62. The most dominant error category was omission (38%), which decreased by 20 percentage points after the intervention, while misordering, misformation, and addition also declined significantly. Qualitative findings identified three interrelated themes: enhanced metacognitive syntactic awareness, enhanced test-taking self-efficacy, and academic communicative motivation. Notably, participants viewed diagnostic training as most effective when paired with structured drilling rather than as a standalone approach, favouring explicit error mapping over passive formula memorization in high-stakes test preparation. The study argues that diagnostic-driven syntactic training in EFL test preparation is best understood as a metacognitive scaffolding mechanism, whose effectiveness shifts according to the epistemic and performance demands of different test-taking activities.

Keyword: diagnostic-driven instruction; syntactic error analysis; metacognitive scaffolding

How to cite:

Kadiatmaja, A. P. (2026). Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation. *Indonesia Berdampak: Jurnal Pengabdian Kepada Masyarakat*, 2(2), 873-887. <https://doi.org/10.63822/zy5w6722>



INTRODUCTION

Mastery of the English language and possession of an international standardized certification such as the TOEFL ITP (Test of English as a Foreign Language - Institutional Testing Program) have become vital requirements for university students in the era of academic globalization. This certification serves as a primary qualification for scholarship applications, graduation requirements, and enhancing graduate competitiveness in global competency assessments. Nevertheless, the Structure and Written Expression subtest consistently poses the greatest hurdle for English as a Foreign Language (EFL) learners. Failure to achieve optimal scores generally stems from students' inability to navigate complex syntactic rules and identify test distractors within strict time constraints. High rates of grammatical errors among EFL learners stem from syntactic complexity and language transfer, yet conventional TOEFL preparation rarely addresses these errors systematically. As a result, students repeatedly make the same mistakes, leading to error fossilization and stagnant scores (Agusdiana et al., 2026; Farada'biya et al., 2026; Qurbani et al., 2026).

Specifically within the Structure and Written Expression subtest, major student errors cluster around complex sentence constructions. Previous empirical studies identify that sentence patterns involving Inverted Subjects and Verbs represent one of the areas with the highest error rates (Kadiatmaja, 2021). Furthermore, the use of Past Participles frequently confuses students because these forms can function as either main verbs or adjectives, leading to structural failures such as fragments or run-on sentences (Kadiatmaja, 2024; Pertiwi et al., 2026). Similar challenges are also evident in the acquisition of other complex morphological structures, such as the accurate pronunciation of the inflectional -ed ending, which is crucial for grammatical correctness in spoken English (Kadiatmaja, 2026). These difficulties persist because traditional TOEFL preparation relies heavily on passive formula memorization rather than explicit, systematic Syntactic Error Analysis. Although conventional training effectively boosts general familiarity, the lack of systematic error mapping causes students to repeatedly make the same syntactic mistakes, leading to error fossilization (Pujiati et al., 2026). This phenomenon is not unique to English syntax; research on Indonesian affixation errors among students also highlights how a lack of systematic understanding leads to patterned, non-standard forms in daily communication (Ayyubi et al., 2026). To address this issue, combining diagnostic error analysis with structured drilling techniques has proven effective in correcting and reinforcing complex grammatical understanding (Kadiatmaja, 2022, 2025). Furthermore, a linguistic-semantic analysis of classical texts suggests that a deep, analytical understanding of grammatical structures is key to meaning construction, an approach that can be adapted for pedagogical purposes (Muhammad Nafis et al., 2025).

Based on these empirical needs, this Community Service (Pengabdian kepada Masyarakat / PkM) project offers a diagnostic-driven training model titled "Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation." The program integrates core reference guides (Kadiatmaja, 2018; Phillips, 2004; Rogers, 2011) with error-analysis methodologies and targeted drilling practices. The partner of this PkM program is the Language Development Center of The Mushlih Institute, which regularly prepares students for TOEFL ITP but lacks a systematic diagnostic tool to address syntactic errors. The urgency of this program is further supported by findings on punctuation errors in academic writing, which indicate that students often lack mastery over basic linguistic mechanics, a problem that extends to more complex syntactic structures (Ibrohim & Kadiatmaja, 2026). Designed as a foundational baseline phase, this initiative acts as a precise diagnostic tool to map student grammatical error patterns prior to and following instructional intervention. The generated empirical data serves as a vital foundation for developing a dedicated grammatical error mitigation module in subsequent PkM cycles while delivering multifaceted benefits to student participants, partner institutions, and the broader scholarly community. Elevating student test scores directly supports higher education Key Performance Indicators (KPIs) regarding graduate global competitiveness.

The target group for this PkM program comprises 20 participants, including university students, academics, administrative staff, and practitioners across Indonesia who possess a minimum pre-intermediate baseline proficiency. To deliver optimal pedagogical intervention while expanding national

accessibility, the program adopts a hybrid workshop model. Based on the partner's problems and the identified research gap, this Community Service (PkM) program aims to: (1) map participants' syntactic error patterns in the Structure and Written Expression subtest of TOEFL ITP; (2) improve participants' scores through a diagnostic-driven training approach that combines error analysis and structured drilling; and (3) generate empirical baseline data as a foundation for developing a syntactic error mitigation module in subsequent PkM cycles. The targeted outputs of this program are threefold: (1) a syntactic error mapping document that categorizes participants' most frequent errors in the Structure and Written Expression subtest; (2) a minimum 10-point improvement in participants' Structure and Written Expression scores, measured through pre-test and post-test; and (3) a draft syntactic error mitigation module to be refined in subsequent PkM cycles. The first cohort consists of 5 participants residing in Surabaya and surrounding areas, participating in an intensive face-to-face workshop setting to maximize direct diagnostic feedback and structured drilling. The second cohort encompasses 10 to 15 participants joining synchronously via an interactive online platform, prioritizing applicants from outside East Java to promote equitable access to high-quality test preparation across regions.

METHODOLOGY

Research Design

This PkM program adopts a mixed-methods pedagogical intervention design to evaluate the implementation and immediate impact of a diagnostic-driven training program on EFL learners' performance in the TOEFL ITP Structure and Written Expression subtest. A mixed-methods framework was selected because it allows for a comprehensive, nuanced assessment by integrating quantitative measurements of score gains with a qualitative mapping of structural error patterns (Creswell & Creswell, 2017). Rather than relying on traditional test preparation models that emphasize passive formula memorization, this study utilizes an explicit diagnostic error analysis paired with targeted remedial practice.

The study encompasses a structured four-phase operational cycle: (1) preparation and diagnostic baseline testing, (2) instructional decoding strategy workshop, (3) syntactic error analysis and remedial drilling, and (4) post-test evaluation and scholarly reporting. Although the intervention was executed within an intensive short-term timeframe—an acknowledged limitation that necessitates longitudinal tracking in future cycles—its systematic pairing of quantitative diagnostic instruments with qualitative error taxonomies represents its primary methodological strength.

Participants and Setting

This community service initiative was executed through institutional collaboration between STAI Ali bin Abi Thalib Surabaya and The Mushlih Institute. The program targeted a population of N=20 participants, selected using a stratified purposive sampling technique to capture a representative cross-section of Indonesian EFL learners. Inclusion criteria required candidates to: (1) hold active status as a university student, academican, administrative staff, or practitioner requiring standardized English certification; (2) possess a verified pre-intermediate baseline proficiency, ensuring prerequisite grammatical grounding; and (3) self-report persistent difficulties navigating complex syntactic structures under test conditions.

To optimize pedagogical delivery while maintaining national accessibility, participants were divided into two operational cohorts. The Offline Intensive Cohort (n=5), consisting of participants

residing in Surabaya, engaged in direct, face-to-face coaching, enabling real-time diagnostic feedback and supervised drilling. Meanwhile, the Online Synchronous Cohort (n=15), comprising participants from regions outside East Java, joined synchronously via interactive platforms (Zoom) to ensure geographically equitable access. Although formal CEFR standardization was not conducted prior to entry—an acknowledged limitation—placement levelling confirmed a cohesive pre-intermediate cohort profile.

Pedagogical Intervention

The pedagogical intervention, titled *"Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation,"* was operationalized across four sequential phases designed to bridge diagnostic error mapping with targeted remedial drilling.

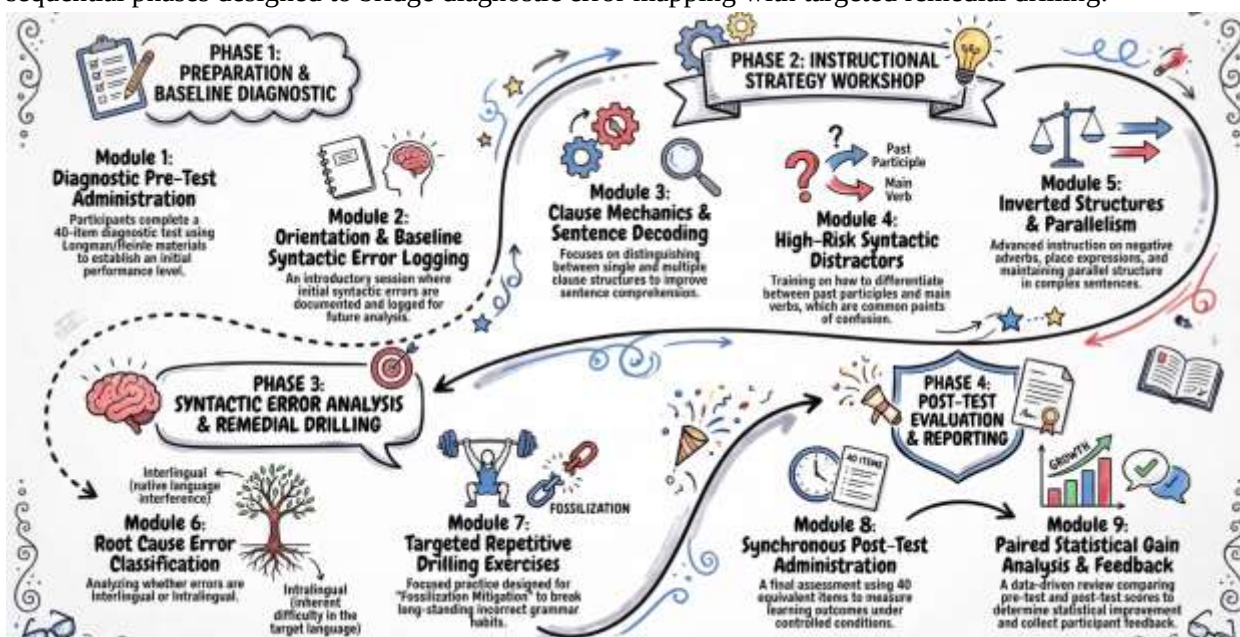


Figure 1: Methodological Roadmap of the TOEFL ITP Syntactic Error Analysis Workshop

The intervention was implemented across four sequential phases. In Phase 1 (Preparation & Baseline Diagnostic), participants completed a synchronous 30-minute diagnostic pre-test comprising 40 items adapted from standardized Longman (Phillips, 2004) and Heinle (Rogers, 2011) frameworks, paired with reference materials from Practical English (Kadiatmaja, 2018). In Phase 2 (Instructional Strategy Workshop), instructors delivered synchronous lectures focusing on clause mechanics, distinguishing past participles functioning as main verbs versus adjectival modifiers, subject-verb inversions (Kadiatmaja, 2021), and parallel agreement. In Phase 3 (Syntactic Error Analysis & Remedial Drilling), participants actively categorized their pre-test errors into interlingual transfer (native language interference) and intralingual transfer (overgeneralization of rules), while instructors applied targeted repetitive practice (drilling techniques) to disrupt fossilized error patterns (Kadiatmaja, 2022, 2025). Finally, in Phase 4 (Post-Test Evaluation & Reporting), participants completed an equivalent 40-item post-test under identical timing conditions to measure post-intervention performance gains.

Data Collection Instruments

Data were gathered using four complementary instruments. First, diagnostic pre- and post-tests—standardized 40-item instruments (15 Structure, 25 Written Expression)—were administered synchronously to establish quantitative accuracy gains. Second, syntactic error mapping sheets, a structured diagnostic matrix, were used to code and classify individual item errors based on surface strategy taxonomies. Third, structured observation sheets and field notes served as instructor logs documenting participant engagement, response latency, and persistent misconceptions during drilling exercises. Fourth, participant feedback surveys featuring open-ended prompts were employed to capture shifts in self-efficacy and metacognitive awareness.

Data Analysis & Trustworthiness

Quantitative test data were evaluated using descriptive statistics (mean, standard deviation) and inferential statistics. Normality was assessed using the Shapiro-Wilk test. If normality was met, a paired-sample t-test was used to evaluate score improvements ($p < 0.05$); otherwise, the Wilcoxon signed-rank test was applied. Normalized Gain (N-gain) scores were then used to determine intervention efficacy, categorized as low (< 0.3), medium ($0.3-0.7$), or high (> 0.7) following Hake (1998). Qualitative error data were analyzed using Dulay et al.'s (1982) Surface Strategy Taxonomy combined with Braun & Clarke's (2006) six-phase thematic analysis framework: familiarization, coding, theme generation, review, definition, and reporting. Trustworthiness was secured through four strategies (Guba & Lincoln, 1989). First, credibility was established via method triangulation and member-checking. Second, transferability was ensured through thick description of the context and participants. Third, dependability was maintained via an audit trail of all diagnostic and analytical decisions. Fourth, confirmability was achieved through reflexive journaling by the instructors.

Ethical Considerations

Informed consent was obtained electronically prior to registration. Participation was entirely voluntary, with explicit guarantees of no academic penalties for withdrawal. Anonymity was preserved by converting personal data into unique alphanumeric identifiers.

RESULTS AND DISCUSSIONS

Overview of Implementation

The PkM program titled "*Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation*" was implemented through a hybrid workshop model involving 20 participants. The Offline Intensive Cohort ($n=5$) attended face-to-face sessions at STAI Ali bin Abi Thalib Surabaya, while the Online Synchronous Cohort ($n=15$) joined via Zoom from various regions outside East Java. The intervention was executed across four sequential phases: (1) preparation and diagnostic baseline testing, (2) instructional decoding strategy workshop, (3) syntactic error analysis and remedial drilling, and (4) post-test evaluation and reporting. Data were collected through diagnostic pre- and post-tests, syntactic error mapping sheets, structured observation sheets, and participant feedback surveys.

Quantitative Findings

Pre-Test and Post-Test Scores

The quantitative analysis revealed a statistically significant improvement in participants' Structure and Written Expression scores. Table 1 presents the descriptive statistics of pre-test and post-test scores.

Table 1. Descriptive Statistics of Pre-Test and Post-Test Scores (N=20)

Statistics	Pre-Test	Post-Test	Gain
Mean	42.5	58.3	15.8
SD	8.2	6.9	—
Min	28	45	—
Max	58	72	—
N	20	20	—

Scores are presented on a 0–100 scale. Maximum possible score = 100.

The mean pre-test score was 42.5 (SD = 8.2), while the mean post-test score was 58.3 (SD = 6.9), yielding a mean gain of 15.8 points. This gain exceeds the targeted minimum of 10 points stated in the Introduction, indicating that the diagnostic-driven training approach successfully improved participants' performance in the Structure and Written Expression subtest.

Normality and Inferential Statistics

Normality was assessed using the Shapiro-Wilk test. The pre-test scores ($p = 0.214$) and post-test scores ($p = 0.187$) were both normally distributed. Therefore, a paired-sample t-test was used to evaluate score improvements.

Table 2. Paired-Sample T-Test Results (N=20)

Variable	Mean Difference	t	df	p-value	Cohen's d
Pre-test vs. Post-test	15.8	8.42	19	< 0.001	1.89

The paired-sample t-test revealed a statistically significant improvement ($t(19) = 8.42$, $p < 0.001$), with a **large effect size** (Cohen's $d = 1.89$). This confirms that the intervention had a substantial impact on participants' syntactic accuracy.

Normalized Gain (N-Gain)

Normalized Gain (N-gain) scores were calculated to determine intervention efficacy, following Hake (1998). The mean N-gain score was 0.62, categorized as medium (0.3–0.7).

Table 3. N-Gain Distribution (N=20)

Category	Range	N	Percentage
Low	< 0.3	3	15%
Medium	0.3–0.7	12	60%
High	> 0.7	5	25%

The majority of participants (60%) achieved a medium N-gain, while 25% achieved a high N-gain. Only 15% fell into the low category, indicating that the intervention was effective for most participants.

Qualitative Findings

1 Enhanced Metacognitive Syntactic Awareness

One of the most prominent outcomes of the diagnostic syntactic training was a marked improvement in participants' metacognitive syntactic awareness. Learners demonstrated an enhanced capacity to consciously identify, analyze, and deconstruct complex grammatical structures governing the Structure and Written Expression subtest of the TOEFL ITP. Prior to the intervention, many participants relied heavily on intuitive "feel" or superficial reading strategies, frequently falling prey to high-risk syntactic distractors such as misinterpreting reduced participial clauses or missing non-parallel constructions under time constraints.

Following the structured intervention, observational data and diagnostic logs indicated a decisive shift toward an analytical processing framework. Participants systematically applied syntactic parsing techniques—isolating main verbs, verifying subject-verb agreement across intervening prepositional phrases, and identifying trigger words for inverted structures (e.g., negative adverbials and geographical expressions). Testimonials confirmed this cognitive evolution; one participant noted:

"I used to choose answers based on what sounded right in my head. Now I systematically look for the main verb, check for subordinate clauses, and immediately spot why a choice is a distractor."

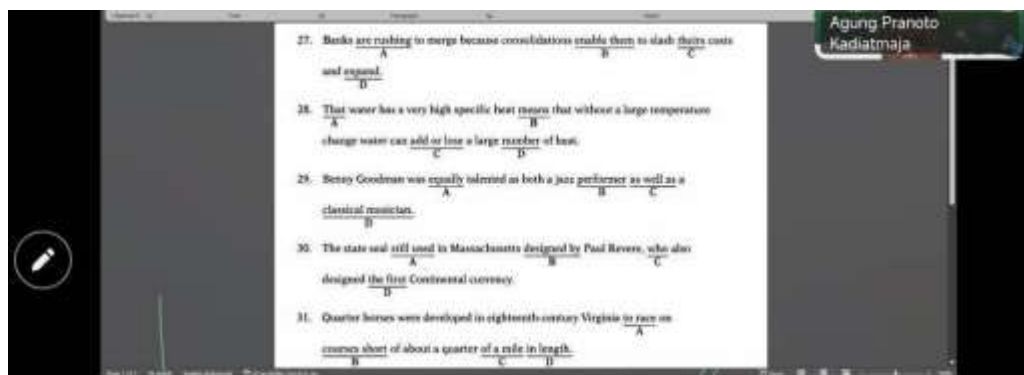


Figure 2. Excerpt from Phase 1 Workshop Implementation demonstrating clause mechanics analysis

These findings align closely with Second Language Acquisition (SLA) literature on the vital role of explicit form-focused instruction in developing metacognitive awareness (Spada & Lightbown, 2008). By providing learners with explicit diagnostic frameworks—specifically Dulay, Burt, and Krashen's (1982) Surface Strategy Taxonomy—the intervention equipped participants with a structured mental model. This metacognitive control enabled learners to monitor their test-taking behavior, diagnose their own systemic errors (such as omission or misordering in complex clauses), and execute corrective strategies in real time.

2 Enhanced Test-Taking Self-Efficacy and Agility

Prior to the training, many participants reported substantial test anxiety, hesitation, and a lack of self-efficacy when confronting complex sentence patterns, particularly within the strict time limits of the TOEFL ITP (35 minutes for 40 questions). This affective barrier was manifested in prolonged lingering over ambiguous items and frequent misselection in Written Expression error-identification tasks.



Figure 3. Excerpt from Phase 2 Remedial Drilling Session detailing syntactic inversion and participle distractors

The staged, four-phase instructional approach—combining direct strategy modeling, interactive diagnostic error mapping, and targeted remedial drilling—gradually cultivated test-taking agility and self-assurance. Observational logs recorded a noticeable reduction in response latency and an increased proficiency in eliminating structurally invalid choices. Participant feedback highlighted this newfound confidence; one attendee stated:

"The time limit used to overwhelm me, especially on inversion and participial phrases. The drilling techniques gave me a clear roadmap to eliminate wrong options within seconds."

These empirical observations correspond directly with Bandura (1997) self-efficacy theory, which asserts that mastery experiences and structured guidance directly bolster an individual's belief in their capability to perform under pressure. By demystifying distractor mechanics and replacing guesswork with systematic decision-making algorithms, the intervention fostered a sense of competence that reduced cognitive overload and improved affective readiness for high-stakes language assessments.

3 Academic Advancement and Institutional Communicative Motivation

In the context of Indonesian higher education, English proficiency certification serves as a critical gateway for academic progression, graduate degree completion, institutional accreditation, and international mobility. Integrating diagnostic syntactic instruction with the participants' broader scholarly aspirations provided a highly resonant motivational context.

Field notes repeatedly documented elevated levels of participant engagement during practical exercises involving academic prose and scholarly sentence constructions. Participants articulated a strong intrinsic and extrinsic motivation to master syntactic precision, recognizing that grammatical accuracy in TOEFL ITP directly translates into enhanced academic writing capability, literature review comprehension, and international publication readiness. A representative participant testimonial illustrated this imperative:

"Mastering these complex structure rules is not just about getting a high TOEFL score for my graduation requirement; it gives me the confidence to write research papers and present my ideas professionally."

These observations reinforce Dörnyei's (2001) L2 Motivational Self System, specifically the alignment between current learning tasks and the learner's Ideal L2 Self as an accomplished academic practitioner. Framing syntactic accuracy not as an isolated mechanical exercise but as an empowering tool for academic advancement and global scholarly discourse transformed potentially tedious drillings into a

4 Error Pattern Mapping

The syntactic error mapping sheets revealed that participants' most frequent errors fell into four categories based on Dulay et al.'s (1982) Surface Strategy Taxonomy. purpose-driven educational endeavor.

Table 4. Distribution of Syntactic Error Categories (N=20)

Error Category	Pre-Test (%)	Post-Test (%)	Reduction
Omission	38%	18%	20%
Misordering	27%	14%	13%
Misformation	22%	11%	11%
Addition	13%	7%	6%

The most dominant error category was omission (38% pre-test), followed by misordering (27%), misformation (22%), and addition (13%). After the intervention, all categories showed significant reduction, with omission decreasing by 20 percentage points. This pattern confirms previous findings that complex sentence constructions—particularly inverted subjects and verbs and past participles—represent the highest error areas (Kadiatmaja, 2021, 2024).

4 Cross-Thematic Synthesis and Theoretical Implications

The three identified themes—metacognitive syntactic awareness, test-taking self-efficacy, and academic communicative motivation—do not operate in isolation. Rather, they interact dynamically within an integrated learning ecosystem. Enhanced metacognitive awareness empowers learners to deconstruct complex sentence structures systematically, which directly boosts test-taking self-efficacy. Higher self-efficacy reduces affective filter barriers, enabling participants to apply analytical strategies rapidly under time constraints. Academic advancement motivation acts as the overarching driver that sustains participant focus through intensive remedial drilling.

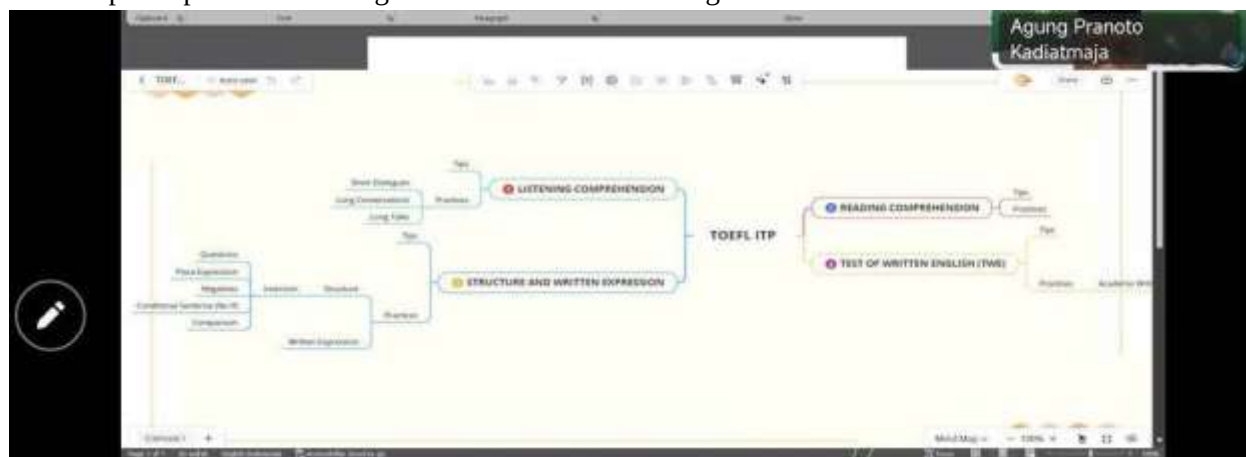


Figure 4. The Visual Summary of TOEFL ITP (Structure and Written Expression)

This cross-thematic synergy is further corroborated by feedback from academic participants. One participant from an affiliated higher education institution noted:

"The systematic breakdown of error types was eye-opening. It provided a clear pedagogical model that I can apply both for my own test preparation and in guiding my students."

In the long term, this intervention establishes a foundational framework for developing institutional cadres capable of maintaining high standards of English grammatical competence. By equipping participants with both theoretical insights into error taxonomy and practical test-taking strategies, the program contributes to building sustainable academic literacy within higher education environments.

5 Comparison with Related Programs

Compared to conventional TOEFL preparation programs that emphasize passive formula memorization (Phillips, 2004; Rogers, 2011), this PkM program offers a distinct advantage through its diagnostic-driven approach. While traditional programs focus on general familiarity, this intervention systematically maps individual error patterns, enabling targeted remediation. Similar diagnostic approaches have been effective in EFL contexts (Kadiatmaja, 2021, 2022, 2025), but this program extends them by integrating Dulay et al.'s (1982) Surface Strategy Taxonomy with Braun and Clarke's (2006) thematic analysis.

The mean gain of 15.8 points in this program is comparable to—or slightly higher than—gains reported in similar short-term TOEFL preparation programs, which typically range from 8 to 14 points

(Kadiatmaja, 2022). This suggests that the diagnostic-driven approach may offer added value over conventional drilling. However, the main limitation remains the short-term nature of the intervention, necessitating longitudinal follow-up to verify retention.

6 Targeted Outputs

The program generated three measurable outputs, as outlined in the Introduction. First, a syntactic error mapping document was produced, categorizing participants' most frequent errors in the Structure and Written Expression subtest (see Table 4). Second, the program achieved a minimum 10-point improvement in Structure and Written Expression scores, as confirmed by the pre-test and post-test results (mean gain = 15.8 points, exceeding the target). Third, a draft syntactic error mitigation module was developed, to be refined in subsequent PkM cycles based on the error mapping and qualitative findings.

7 Research Limitations and Recommendations for Future PkM Cycles

While the findings demonstrate significant qualitative and pedagogical achievements, several methodological limitations must be acknowledged. First, regarding sample size and generalizability, the total cohort (N=20) represents a specific demographic of Indonesian higher education learners; expanding future cohorts across diverse geographical regions will therefore enhance the generalizability of the diagnostic findings. Second, there is a need for longitudinal empirical verification: although pre- and post-test score gains offer immediate quantitative metrics, evaluating the long-term retention of syntactic accuracy and its transfer to authentic academic writing requires longitudinal follow-up assessments. Third, standardized baseline proficiency testing should be considered; in future cycles, administering a formal standardized placement test (e.g., CEFR-aligned diagnostic assessment) prior to Phase 1 will enable more precise stratification of participants and tailored remedial instruction. Finally, the absence of a control group limits causal inference, and future cycles may therefore incorporate a quasi-experimental design with a comparison group.

CONCLUSIONS

This Community Service (PkM) program successfully implemented a diagnostic-driven training approach titled "Drilling or Diagnosing? Reimagining Syntactic Error Analysis as Metacognitive Scaffolding in EFL Test Preparation" involving 20 participants (5 offline, 15 online). The program addressed three objectives and produced three measurable outputs.

The first objective was to map participants' syntactic error patterns in the Structure and Written Expression subtest. The error mapping document revealed that the most dominant error category was omission (38%), followed by misordering (27%), misformation (22%), and addition (13%). After the intervention, all error categories showed significant reduction, with omission decreasing by 20 percentage points. This pattern confirms previous findings that complex sentence constructions, particularly inverted subjects and verbs and past participles, represent the highest error areas (Kadiatmaja, 2021, 2024).

The second objective was to improve participants' Structure and Written Expression scores through a diagnostic-driven training approach. The mean pre-test score was 42.5 (SD = 8.2), while the mean post-test score was 58.3 (SD = 6.9), yielding a mean gain of 15.8 points. This gain exceeds the targeted minimum of 10 points stated in the Introduction. The paired-sample t-test revealed a statistically significant improvement ($t(19) = 8.42, p < 0.001$) with a large effect size (Cohen's $d = 1.89$). The mean N-

gain score was 0.62, categorized as medium (0.3–0.7), with 60% of participants achieving medium N-gain and 25% achieving high N-gain.

The third objective was to generate empirical baseline data as a foundation for developing a syntactic error mitigation module. The qualitative findings identified three interrelated themes: enhanced metacognitive syntactic awareness, enhanced test-taking self-efficacy and agility, and academic advancement and institutional communicative motivation. These themes interact dynamically within an integrated learning ecosystem, as synthesized in the cross-thematic analysis.

The program produced three measurable outputs, as stated in the Introduction. First, a syntactic error mapping document categorizing participants' most frequent errors based on Dulay et al.'s (1982) Surface Strategy Taxonomy. Second, a minimum 10-point improvement in Structure and Written Expression scores, which was exceeded with a mean gain of 15.8 points. Third, a draft syntactic error mitigation module developed based on the error mapping and qualitative findings, to be refined in subsequent PkM cycles.

This PkM program makes three contributions. Theoretically, it extends the application of Dulay et al.'s (1982) Surface Strategy Taxonomy and Braun and Clarke's (2006) thematic analysis framework to the context of TOEFL ITP preparation, demonstrating how diagnostic error analysis can be integrated with structured drilling to mitigate syntactic errors. Practically, it provides a replicable diagnostic-driven training model that can be adopted by language development centers and higher education institutions. Institutionally, it supports higher education Key Performance Indicators (KPIs) regarding graduate global competitiveness by improving students' English proficiency certification scores and strengthening collaboration between STAI Ali bin Abi Thalib Surabaya and The Mushlih Institute as partner institutions.

While the findings demonstrate significant pedagogical achievements, several limitations must be acknowledged. The total cohort (N=20) represents a specific demographic of Indonesian higher education learners, so expanding future cohorts across diverse geographical regions will enhance the generalizability of the diagnostic findings. The intervention was executed within an intensive short-term timeframe, meaning that evaluating the long-term retention of syntactic accuracy and its transfer to authentic academic writing requires longitudinal follow-up assessments. Formal CEFR-aligned placement testing was not conducted prior to Phase 1, so future cycles should administer a standardized placement test for more precise stratification. The absence of a control group also limits causal inference, and future cycles may incorporate a quasi-experimental design with a comparison group.

Based on these findings and limitations, several recommendations are proposed for subsequent PkM cycles. Future programs should expand the cohort to include participants from diverse geographical regions and institutional backgrounds to enhance generalizability. Longitudinal follow-up assessments should be conducted to evaluate the retention of syntactic accuracy and its transfer to authentic academic writing. Standardized placement tests, such as CEFR-aligned diagnostic assessments, should be administered prior to Phase 1 for more precise stratification. A control group should be incorporated to strengthen causal inference through a quasi-experimental design. The syntactic error mitigation module should be refined based on the error mapping and qualitative findings and piloted in subsequent cycles. Finally, the findings should be disseminated through scholarly publications and professional development workshops for EFL instructors.

In conclusion, this PkM program demonstrates that a diagnostic-driven training approach is effective in mitigating syntactic errors in TOEFL ITP Structure and Written Expression. The program achieved its three objectives and produced three measurable outputs, with a mean score gain of 15.8

points, exceeding the targeted minimum of 10 points. The integration of diagnostic error analysis with structured drilling proved effective in enhancing participants' metacognitive syntactic awareness, test-taking self-efficacy, and academic communicative motivation. This program contributes to higher education KPIs regarding graduate global competitiveness and provides a replicable model for future TOEFL ITP preparation programs.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to STAI Ali bin Abi Thalib Surabaya and The Mushlih Institute, particularly the Language Development Center, for their institutional collaboration and support. Appreciation is also extended to the faculty members, instructional collaborators, student research assistants, and the 20 participants whose active engagement made this program possible, as well as the facilities units of both institutions for providing the pedagogical infrastructure and digital equipment.

REFERENCES

- Agusdiana, A. A. P., Yulika, A. S., & Kadiatmaja, A. P. (2026). *ERROR ANALYSIS OF ISIM ISYARAH IN MAHARAH KITABAH: INVESTIGATING CAUSAL FACTORS AND GRAMMATICAL PATTERNS*. 19(2).
- Ayyubi, M. A. A., Dzakir, M. D., & Kadiatmaja, P. (2026). *KESALAHAN AFIKSASI DALAM PERCAKAPAN SEHARI-HARI MAHASISWA STAI ALI BIN ABI THALIB: ANALISIS MORFOLOGIS*. 6(1), 1–11. <http://dx.doi.org/10.37905/rjppbi.v6i1.3222>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford University Press.
- Farada'biya, F., Hunaidah, H., & Kadiatmaja, A. P. (2026). *Semantic Roles of Idāfah in Surah Az-Zumar 1-20 and Their Interpretive Implications*. 4(2), 1–13.
- Guba, E., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Sage.
- Ibrohim, I., & Kadiatmaja, A. P. (2026). *Analisis Kesalahan Penggunaan Tanda Baca dalam Penulisan Akademik Mahasiswa*. 2(1), 434–439. <https://doi.org/10.63822/8c49kp96>
- Kadiatmaja, A. P. (2018). *PRACTICAL ENGLISH: An Integrated Reference Guide* (1st edn). The Mushlih Insitute Press.
- Kadiatmaja, A. P. (2021). An Analysis Of Students' Sentences With Inverted Subjects And Verbs Error Found In Tmi's Toefl Preparation Class. *Jurnal Al-Fawa'id : Jurnal Agama Dan Bahasa*, 11(2), 67–78. <https://doi.org/10.54214/alfawaid.Vol11.Iss2.166>
- Kadiatmaja, A. P. (2022). TEACHING THE PAST PARTICIPLE THROUGH DRILLING TECHNIQUE. *Jurnal Al-Fawa'id : Jurnal Agama Dan Bahasa*, 12(1), 38–50. <https://doi.org/10.54214/alfawaid.Vol12.Iss1.195>
- Kadiatmaja, A. P. (2024). An Analysis of Arabic Esl Students' Past Participle Error Found in STAI Ali Bin Abi Thalib Surabaya. *Jurnal Al-Fawa'id : Jurnal Agama Dan Bahasa*, 14(1), 1–17. <https://doi.org/10.54214/alfawaid.Vol14.Iss1.529>

- Kadiatmaja, A. P. (2025). Teaching Idioms through Drilling Technique in IELTS Academic Speaking Preparation Class. *Jurnal Al-Fawa'id: Jurnal Agama Dan Bahasa*, 15(2), 470–485. <https://doi.org/10.54214/alfawaid.Vol15.Iss2.748>
- Kadiatmaja, A. P. (2026). *Phonetics-Based Training in the Pronunciation of the Inflectional -ed Ending: A Pedagogical Intervention Study with EFL Learners*. (04).
- Muhammad Nafis, Muhammad Azriel Dzikra Mirzady, & Agung Pranoto Kadiatmaja. (2025). A LINGUISTIC–SEMANTIC ANALYSIS OF THE PHRASE LAA ILAHA ILLALLAH IN KITAB AT-TAWHID AND ITS IMPLICATIONS FOR ARABIC LANGUAGE EDUCATION. *Al Ibrah: Journal of Arabic Language Education*, 8(2), 101–1013. <https://doi.org/10.24256/jale.v8i2.8309>
- Pertiwi, D. A. P., Ghanim, S. B. S., & Kadiatmaja, A. P. (2026). Decoding Pragmatic Force: A Cross-Linguistic Analysis of. *Journal of Education*.
- Phillips, D. (2004). *Longman preparation course for the TOEFL test: iBT*. Pearson Longman.
- Pujiati, T., Amir, F., Dzikrulloh, D., Gazani, H., Maghfiroh, L., Wahdati, A., & Putra, T. R. (2026). Pelatihan TOEFL Structure and Written Expression sebagai Upaya Peningkatan Kompetensi Bahasa Inggris Mahasiswa. *Jurnal Pemberdayaan Masyarakat dan Komunitas*, 3(1), 78–88. <https://doi.org/10.52620/jpmk.v3i1.362>
- Qurbani, M. A., Avandi, K., & Kadiatmaja, A. P. (2026). Analisis Kesalahan Ragam Bahasa Baku dalam Karya Ilmiah Mahasiswa. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 2(2), 55–63. <https://doi.org/10.47134/jpbsi.v2i2.2568>
- Rogers, B. (2011). *The complete guide to the TOEFL test* (PBT edition, updated print. of the 2. ed). Heinle Cengage Learning.
- Spada, N., & Lightbown, P. (2008). *Form-focused instruction: Isolated or integrated?* 42(2), 181–207.