

## **EVALUATING TOEIC INSTRUCTION THROUGH THE CIPP MODEL: INSIGHTS FROM AN INDONESIAN AVIATION ACADEMY**

**<sup>1</sup>Hendra Sudarso, <sup>1</sup>Safitri Era Globalisasi, <sup>1</sup>Dimas Hari Cahyo, <sup>1</sup>Yasyfa' Kawakibi, <sup>2</sup>Haris Dibdyaningsih**

<sup>1</sup>Akademi Penerbang Indonesia Banyuwangi, Indonesia

<sup>2</sup>STKIP Al Hikmah Surabaya, Indonesia

\*Corresponding author e-mail: [hendrasudarso27@gmail.com](mailto:hendrasudarso27@gmail.com)

### **ABSTRACT**

English proficiency is essential for safe pilot-air traffic controller communication in aviation operations. While many Indonesian flight academies adopt TOEIC as an institutional benchmark for English certification, questions remain about its effectiveness in measuring the communicative competence required for operational aviation contexts. This qualitative study evaluated TOEIC instruction at an Indonesian Flight Academy using the CIPP (Context, Input, Process, Product) evaluation model. Data revealed a significant gap between strong institutional policy endorsement of TOEIC and actual classroom practice. Although institutional frameworks mandated TOEIC certification, curriculum implementation remained insufficiently aligned with ICAO language proficiency standards, instructors lacked specialized aviation English training, and teaching focused predominantly on test preparation rather than authentic communication scenarios. Notably, while cadets achieved respectable TOEIC scores (550–780), their operational communicative readiness during simulated aviation tasks remained limited. The study recommends curriculum realignment linking TOEIC objectives with ICAO descriptors, mandatory ESP professional development for instructors, and integration of simulation-based communicative tasks. Learner self-efficacy emerged as a crucial psychological factor mediating the transfer of language learning to real communication. This evaluation contributes to ESP pedagogy and aviation English policy by demonstrating how institutional, pedagogical, and psychological dimensions interact to shape language instruction effectiveness in professional, safety-critical contexts.

### **ARTICLE INFO**

#### **Article History:**

*Received: 23 October 2025*

*1<sup>st</sup> revision: 17 November 2025*

*2<sup>nd</sup> revision: 17 December 2025*

*Accepted: 27 December 2025*

*Published: 30 December 2025*

#### **Keywords:**

*TOEIC instruction;*

*CIPP model;*

*Aviation English;*

*ESP evaluation;*

*ICAO language proficiency;*

*Self-efficacy*

**How to cite:** Sudarso, H., Globalisasi, S. E., Cahyo, D. H., Kawakibi, Y., & Dibdyaningsih, H. (2025). Evaluating TOEIC Instruction through the CIPP Model: Insights from an Indonesian Aviation Academy. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP*, 12(2), 614–625. <https://doi.org/10.33394/jo-elt.v12i2.18146>

Copyright© 2025, Hendra Sudarso, Safitri Era Globalisasi, Dimas Hari Cahyo, Yasyfa' Kawakibi, Haris Dibdyaningsih

This is an open access article under the [CC BY-SA](#) License.



## **INTRODUCTION**

English functions as the operational lingua franca of international aviation, enabling precise and effective communication between pilots and air traffic controllers (ATCs) in both routine and non-routine situations. Because aviation communication directly affects

operational safety, the International Civil Aviation Organization (ICAO) has established strict language proficiency requirements to minimize misunderstandings and reduce the risk of accidents (Dewi & Ohi, 2023; Read & Knoch, 2009; Garcia & Fox, 2020). These requirements emphasize not only standardized phraseology but also the ability to communicate clearly, accurately, and flexibly under time pressure.

ICAO mandates that pilots and ATCs demonstrate adequate English proficiency as an integral component of operational competence, supported by a rating scale ranging from Level 1 (Pre-Elementary) to Level 6 (Expert), with Level 4 designated as the minimum operational standard (Masiulionienė & Tupčiauskaitė, 2023; Bystrova et al., 2019). Research has repeatedly shown that inadequate language proficiency and communicative breakdowns have contributed to aviation incidents and accidents, highlighting the safety-critical nature of aviation English competence (Alderson, 2009; Abeyratne, 2008; Coertze et al., 2014; Clark & Williams, 2020; Dewi & Ohi, 2023).

In Indonesia, many aviation academies rely on the Test of English for International Communication (TOEIC) as an institutional benchmark for assessing English proficiency among future aviation professionals (Park, 2018; Huang, 2013). TOEIC is widely adopted due to its practicality, standardization, and international recognition. However, its primary focus on general workplace English has raised concerns regarding its relevance to aviation contexts, which require specialized vocabulary, interactional competence, and the ability to manage communication in high-pressure operational settings (Kim, 2018; Demirdöken, 2024). Scholars argue that TOEIC predominantly assesses receptive skills and controlled language use, offering limited insight into the interactive and dynamic communication essential for aviation safety (Farris et al., 2008; Mathews et al., 2022).

Empirical evidence indicates that TOEIC-oriented instruction often emphasizes drill-based test preparation, which may improve test familiarity and general listening skills but remains disconnected from aviation-specific discourse and rapid decision-making demands (García, 2023; Fauzi, 2020). Consequently, learners may achieve satisfactory TOEIC scores while experiencing difficulty in operational communication, particularly in non-routine or emergency scenarios (Vitryak et al., 2016). This discrepancy suggests a pedagogical and evaluative gap between standardized proficiency measurement and the communicative competence required by ICAO standards (Arslan, 2021; Alderson, 2010).

Although previous studies have examined TOEIC washback and aviation English instruction, much of the existing research focuses on general vocational contexts or evaluates outcomes solely through test scores. Limited attention has been given to systematic program evaluation that examines how institutional policy, instructional resources, pedagogical processes, and learning outcomes interact within aviation education (Kim & Elder, 2014; Zhao et al., 2017; Bystrova et al., 2019; Oleksandrivna & Kharlamova, 2019). Moreover, few studies investigate whether TOEIC-based instruction adequately prepares learners for high-stakes operational communication, where communicative failure has direct safety implications (Kraśnicka, 2016; Torres, 2023).

To address this gap, the present study evaluates TOEIC instruction in an Indonesian aviation academy using the CIPP (Context, Input, Process, Product) evaluation model. This framework enables a comprehensive analysis of instructional effectiveness by examining policy alignment, resource adequacy, pedagogical practices, and learning outcomes in an integrated manner. By moving beyond test-score analysis, the study seeks to determine whether TOEIC instruction supports the development of communicative competence aligned with ICAO operational requirements.

This study aims to (1) examine the contextual alignment between TOEIC-based instruction and ICAO language proficiency standards, (2) evaluate instructional inputs and teaching processes supporting TOEIC preparation, and (3) assess the extent to which TOEIC

instruction contributes to students' operational communicative readiness. The findings are expected to contribute theoretically by strengthening understanding of construct validity and domain authenticity in aviation English assessment, and practically by providing evidence-based recommendations for aligning aviation English curricula with safety-driven communicative demands rather than relying solely on general proficiency testing.

## RESEARCH METHOD

### Research Design

This study employed a qualitative descriptive research design to evaluate the Teaching of English for International Communication (TOEIC) instruction in an Indonesian aviation academy. Guided by the Context, Input, Process, and Product (CIPP) evaluation framework, the study provided a structured yet flexible examination of program implementation, encompassing institutional alignment, instructional conditions, pedagogical practices, and learning outcomes. The CIPP model, originally developed by Stufflebeam, is widely recognized as a robust framework for educational program evaluation across diverse instructional contexts and institutional settings (Setyoko et al., 2016; Saputro et al., 2023; Susanto & Hamzah, 2022).

### Participants

Participants were selected through purposive sampling and consisted of 16 individuals from an Indonesian aviation academy, including 12 student cadets and 4 English instructors involved in TOEIC instruction. Student participants were drawn from the Pilot Simulation Track (PST) and Operational Preparation Unit (OPU) and met the criteria of active enrollment in TOEIC courses, completion of at least one TOEIC mock test, and representation across varied proficiency levels. The instructor participants had a minimum of two years of teaching experience in vocational or aviation-related English programs. In addition to these participants, brief interviews were conducted with two institutional administrators as policy-level informants to provide contextual information related to TOEIC implementation; their input was used exclusively to inform the Context component of the CIPP analysis. All participants provided informed consent, and confidentiality was maintained through the use of pseudonyms.

### Instruments

This study used three qualitative research instruments aligned with the CIPP framework: document analysis guidelines, semi-structured interview protocols, and classroom observation checklists. Documents included institutional policies, curricula, and instructional materials related to TOEIC instruction. Interviews with instructors and students were guided by open-ended questions organized around the CIPP dimensions to capture perceptions and instructional experiences. Classroom observations documented teaching practices, interaction patterns, and the extent of communicative and aviation-specific activities.

### Data Analysis

Data analysis followed the interactive qualitative analysis model proposed by Miles, Huberman, and Saldaña (2020). The process consisted of three iterative stages: data reduction, data display, and conclusion drawing/verification.

First, data reduction involved transcribing interview recordings verbatim and systematically reviewing documents and observation notes. The data were then coded inductively and deductively according to the CIPP dimensions (Context, Input, Process, Product), allowing relevant patterns and themes to emerge while maintaining alignment with the evaluation framework.

Second, data display was conducted by organizing coded data into matrices and thematic tables that facilitated comparison across data sources and participant groups. This step enabled

the identification of consistencies and divergences in instructional practices, institutional alignment, and learning outcomes.

Finally, conclusions were drawn through pattern interpretation and continuously verified through data triangulation across interviews, observations, and documents. This iterative process ensured analytical rigor and enhanced the credibility of the findings.

## RESEARCH FINDINGS AND DISCUSSION

### Research Findings

This study examined TOEIC instruction through classroom observations (four sessions totaling 360 minutes), semi-structured interviews with 12 students and 4 instructors, and document analysis. Two instructional programs were compared: the Pilot Simulation Track (PST), which integrates simulator-based communicative practice into TOEIC instruction, and the Operational Preparation Unit (OPU), which relies primarily on textbook-based test preparation. This comparison enabled an examination of how variations in instructional authenticity, embedded within the research design, relate to learners' communicative readiness and instructional outcomes within the same institutional context.

### Contextual Alignment between TOEIC-Based Instruction and ICAO Language Proficiency Standards

Analysis of institutional documents, administrative interviews, and instructional records revealed a limited contextual alignment between TOEIC-based instruction and ICAO language proficiency standards. Although TOEIC certification was formally mandated as an institutional requirement, no explicit curricular or policy-level integration with ICAO operational language descriptors was identified.

Document analysis showed that TOEIC functioned primarily as a graduation benchmark, with a minimum score requirement specified in institutional regulations. However, course syllabi and curriculum guidelines contained no reference to ICAO Language Proficiency Requirements, the ICAO rating scale, or operational communicative competencies. TOEIC objectives were articulated in terms of score attainment and test readiness rather than aviation-specific communication outcomes.

This lack of alignment was further confirmed through administrative interviews. One administrator explicitly acknowledged the absence of formal integration, stating:

“TOEIC is required because it is widely recognized and easy to administer, but we have not formally aligned TOEIC teaching with ICAO language standards.” (Administrator 1)

Another institutional stakeholder emphasized TOEIC's role as a general indicator rather than an operational measure:

“We consider TOEIC as proof that cadets have English ability, but it does not specifically measure how they communicate in aviation situations.” (Administrator 2)

These statements indicate that TOEIC operates as a proxy measure of general English proficiency, rather than as an assessment grounded in the communicative demands defined by ICAO.

Interviews with instructors further illustrated this contextual gap. Several instructors reported that ICAO standards were not explicitly referenced in instructional planning or course objectives. One instructor noted:

“We focus on helping students achieve the TOEIC score target. ICAO descriptors are not part of our syllabus, so they are rarely discussed in class.” (Instructor 1)

Another instructor similarly stated:

“There is no guideline that connects TOEIC materials with ICAO communication requirements. We rely on the textbook and test format.”  
(Instructor 3)

The absence of formal policy guidance resulted in inconsistent contextualization across instructional programs. While instructors in the Pilot Simulation Track (PST) occasionally incorporated aviation scenarios to contextualize TOEIC content, these efforts were informal and instructor-initiated, rather than institutionally mandated. As one PST instructor explained:

“We try to relate TOEIC listening to pilot–ATC situations, but this is our own initiative, not something required by the curriculum.” (Instructor 2)

In contrast, instructors in the Operational Preparation Unit (OPU) reported no expectation to link TOEIC instruction to ICAO standards:

“In OPU, TOEIC is treated purely as a test. We don’t connect it to ICAO communication because that’s not part of the course objective.” (Instructor 4)

These findings demonstrate that at the Context dimension of the CIPP framework, TOEIC-based instruction was institutionally endorsed without explicit alignment to ICAO language proficiency standards. The absence of documented policy integration, combined with reliance on instructor discretion, resulted in fragmented and uneven contextual alignment. This contextual gap established structural constraints that shaped subsequent instructional inputs and pedagogical processes.

### **Instructional Inputs and Teaching Processes Supporting TOEIC Preparation**

Evaluation of the Input and Process dimensions revealed that TOEIC preparation was supported by adequate institutional resources, yet constrained by limited ESP expertise and predominantly test-oriented pedagogical practices. Although facilities and instructional time were available, the quality and specificity of instructional inputs and teaching processes varied considerably across programs.

Analysis of instructor profiles and interview data indicated that most instructors teaching TOEIC possessed backgrounds in general English rather than aviation English or ESP. Only a small proportion reported having received formal training related to aviation English or ICAO language proficiency standards. One instructor stated:

“My background is general English. I haven’t received specific training in aviation English or ICAO communication.” (Instructor 3)

Another instructor similarly acknowledged:

“We are assigned to teach TOEIC, but there is no requirement for aviation English certification or ESP training.” (Instructor 1)

Instructional materials further reflected this limitation. All observed classes relied exclusively on commercial TOEIC textbooks, with no documented adaptation to aviation contexts. Syllabus and material analysis confirmed the absence of aviation-specific vocabulary, pilot–ATC discourse, or ICAO communicative functions. As one instructor explained:

“We use the TOEIC textbook as it is. There are no aviation scenarios included in the materials.” (Instructor 4)

Despite these constraints, the institution was equipped with language laboratories and flight simulators. However, these facilities were not systematically integrated into TOEIC instruction. Several instructors noted that simulators were considered part of technical flight training rather than language instruction:

“The simulator is mainly used for flying practice, not for English classes.”  
(Instructor 2)

Classroom observations across four sessions (360 minutes) revealed that teaching processes were strongly test-oriented, with approximately 80–85% of instructional time

devoted to listening and reading drills, practice tests, and item analysis. Instruction focused on accuracy, speed, and test-taking strategies rather than communicative production. One student described this emphasis:

“Most of the class is about doing TOEIC questions and checking answers.”  
(Student OPU-3)

Opportunities for speaking practice and interactive communication were limited and largely unassessed. When speaking activities occurred, they were informal and not systematically linked to course objectives. A student commented:

“We rarely practice speaking seriously. If we speak, it’s not part of the TOEIC score.” (Student OPU-5)

Comparative analysis revealed notable differences between the two instructional tracks. In the Pilot Simulation Track (PST), instructors occasionally contextualized TOEIC listening materials by relating them to pilot-ATC communication or simulator scenarios. PST students reported higher engagement and perceived relevance:

“In PST, the teacher sometimes explains TOEIC listening using aviation situations, so it feels more realistic.” (Student PST-2)

However, this contextualization was instructor-initiated rather than institutionally mandated, and its implementation varied. In contrast, instruction in the Operational Preparation Unit (OPU) remained strictly textbook-based, with no integration of aviation scenarios:

“In OPU, TOEIC is treated purely as a test. We focus on finishing the book and practice tests.” (Instructor 4)

The findings indicate that while instructional inputs such as facilities and allocated instructional time were sufficient, key qualitative inputs, particularly ESP-trained instructors and aviation-specific materials, were limited. Teaching processes were dominated by test preparation practices, resulting in minimal communicative production and weak alignment with aviation communication demands. Contextualized instruction occurred only in isolated cases and depended largely on instructor initiative rather than institutional design.

## Learning Outcomes and Communicative Readiness

Evaluation of the Product dimension focused on the outcomes of TOEIC-based instruction in terms of test achievement and operational communicative readiness. The findings indicate that while most cadets met institutional TOEIC score requirements, test performance did not consistently correspond with readiness for operational aviation communication, particularly in non-routine situations.

Document analysis of TOEIC score reports showed that the majority of students achieved the institutional minimum requirement. TOEIC scores ranged from 550 to 780, with a mean score indicative of intermediate to upper-intermediate proficiency. From an institutional perspective, this outcome was regarded as evidence of program effectiveness. One instructor noted:

“Most students pass the TOEIC requirement, so officially the program is considered successful.” (Instructor 1)

Students also perceived TOEIC achievement as the primary indicator of success:

“As long as we reach the TOEIC target score, we are considered ready to graduate.” (Student OPU-4)

These findings suggest that TOEIC scores functioned as the dominant, and in some cases sole, criterion for evaluating learning outcomes.

Despite satisfactory TOEIC scores, classroom observations and simulation data revealed uneven operational communicative performance, particularly during non-routine scenarios. Several cadets with relatively high TOEIC scores demonstrated difficulty sustaining

interaction, negotiating meaning, or responding flexibly when faced with unexpected communication demands. One observation noted:

“A student with a TOEIC score above 700 hesitated and failed to respond appropriately during an unexpected communication breakdown in a simulated task.” (Observation notes)

In contrast, some students with moderate TOEIC scores displayed stronger communicative adaptability, particularly those from the PST program. An instructor observed:

“Some students with lower TOEIC scores perform better in simulations because they are more used to speaking and responding under pressure.” (Instructor 2)

This divergence highlights a validity gap between standardized test outcomes and operational communication performance.

Student interviews further revealed differences in perceived communicative readiness. PST students generally reported higher confidence and greater preparedness for real aviation communication:

“Because we practice communication in simulation, I feel more confident speaking English in aviation situations.” (Student PST-5)

In contrast, many OPU students expressed uncertainty and anxiety when required to communicate spontaneously:

“I know how to answer TOEIC questions, but when I have to speak without text, I feel nervous and confused.” (Student OPU-6)

These perceptions suggest that TOEIC-based instruction supported test readiness more effectively than operational communicative confidence.

The product-level findings indicate that TOEIC-based instruction successfully produced measurable test outcomes aligned with institutional requirements. However, these outcomes did not uniformly translate into communicative readiness for aviation operations, especially in high-pressure or non-routine contexts. Students' operational performance and confidence were more strongly associated with exposure to contextualized and interactive practice than with TOEIC score attainment alone.

These findings underscore the limitations of relying solely on standardized test results as indicators of communicative competence in aviation contexts and highlight the need for complementary performance-based evaluation aligned with ICAO standards.

## Discussion

This study examined the implementation of TOEIC-based instruction within an Indonesian aviation academy using the Context, Input, Process, and Product (CIPP) evaluation framework to assess its alignment with International Civil Aviation Organization (ICAO) language proficiency standards. The findings reveal a systemic misalignment across all CIPP dimensions, indicating that while TOEIC functions effectively as a standardized certification instrument, it remains insufficient in supporting the operational communicative competence required in aviation contexts.

At the contextual dimension, TOEIC is mandated as an institutional benchmark without explicit integration of ICAO language proficiency descriptors. Although ICAO emphasizes interactional competence, clarity under pressure, and the management of non-routine communication, these competencies are not reflected in institutional curriculum objectives or policy frameworks (Fatmawati et al., 2023; Yin et al., 2023; Alderson, 2009). This finding reinforces broader critiques that general-purpose language tests such as TOEIC are prone to construct underrepresentation when applied to specialized and safety-critical domains like aviation (Guedes & Coneglian, 2019; Alderson, 2009). The absence of a formalized connection between TOEIC objectives and ICAO communicative requirements suggests that TOEIC is

perceived primarily as a general proficiency benchmark rather than as a targeted assessment of aviation communication competence (Park, 2018; Torres, 2023).

These contextual ambiguities extend directly into the input and process dimensions. Instructors' expertise in English for Specific Purposes (ESP) is limited, and instructional content relies predominantly on commercial TOEIC textbooks that lack aviation-specific discourse and communicative functions (Masrul & Rasyidah, 2023; Ulfah & Mukminati, 2023; Masiulionienė & Tupčiauskaitė, 2023). Although institutional resources such as simulators are available and hold strong potential for communicative training, they remain underutilized in language instruction, reflecting a persistent disconnect between technical training and language development (Aranda & Henneberry, 2025; Hautasaari & Yamashita, 2015).

Teaching practices are largely test-oriented, prioritizing listening and reading drills aimed at maximizing TOEIC scores rather than fostering communicative competence. Such practices exemplify negative washback effects commonly associated with high-stakes testing, where instructional content narrows to test preparation at the expense of authentic language use (Wilson, 1989; Masrul & Rasyidah, 2023). Nevertheless, variation across instructional tracks indicates that pedagogical authenticity can emerge through instructor agency. In particular, the Pilot Simulation Track demonstrates how localized instructional initiatives may partially mitigate systemic constraints by embedding communicative practice into TOEIC preparation (Fatmawati et al., 2023; Masrul & Rasyidah, 2023). However, these practices remain isolated and are not institutionally standardized.

At the product dimension, a pronounced discrepancy emerges between TOEIC achievement and communicative readiness for operational roles. Although most cadets meet institutional TOEIC requirements, their scores do not consistently correspond with performance in operational communication tasks, especially under high-pressure or non-routine conditions (Fatmawati et al., 2023; Aranda & Henneberry, 2025). Notably, some cadets demonstrate strong operational communication skills despite moderate TOEIC scores, underscoring the limited predictive validity of standardized proficiency tests for real-world professional performance (Alderson, 2009; Fatmawati et al., 2023; Ulfah & Mukminati, 2023). This finding aligns with ICAO's position that operational language use, rather than decontextualized linguistic knowledge, should be the primary indicator of communicative competence in aviation settings (Torres, 2023; Aranda & Henneberry, 2025).

Collectively, these findings demonstrate that TOEIC-based instruction within the examined aviation academy prioritizes certification outcomes over communicative authenticity and operational preparedness. While TOEIC remains a pragmatic and internationally recognized measure of general English proficiency, reliance on it as a sole indicator risks masking deficiencies in interactional competence that are critical for aviation safety and operational effectiveness (Alderson, 2009). By applying the CIPP framework, this study extends prior evaluation research by illustrating how contextual misalignment systematically constrains instructional inputs and teaching processes, ultimately compromising the validity of learning outcomes (Kanno, 2020; Fatmawati et al., 2023).

To address these misalignments, the findings support calls for stronger alignment between institutional language policy and ICAO standards, the enhancement of ESP-focused professional development for instructors, and the adoption of hybrid assessment models that combine standardized testing with performance-based evaluations grounded in authentic aviation communication scenarios. Such measures reaffirm that communicative authenticity is not merely a pedagogical consideration but a fundamental safety imperative in aviation education.

## CONCLUSION

This study evaluated TOEIC-based instruction at the Indonesian Flight Academy using an integrated CIPP–Self-Efficacy framework and found that, despite strong institutional commitment to standardized English proficiency certification, the current instructional model remains misaligned with the communicative demands of aviation operations. The findings demonstrate a systemic cascade in which policy–practice gaps at the contextual level generate deficiencies in instructional inputs and constrain pedagogical processes, ultimately producing a validity gap between TOEIC achievement and operational communicative readiness.

By incorporating learner self-efficacy into program evaluation, this study extends the CIPP framework and contributes to ESP evaluation theory by empirically illustrating how instructional authenticity mediates both communicative performance and learner resilience in high-pressure contexts. The results confirm that standardized proficiency scores alone are insufficient indicators of readiness in safety-critical domains and underscore the necessity of evaluating language competence through operationally grounded performance.

From a practical perspective, the study highlights the urgent need for ICAO-aligned curriculum redesign, sustained ESP-focused professional development, systematic integration of simulation-based communicative tasks, and the adoption of hybrid assessment models that combine standardized testing with performance-based evaluation. Such reforms are essential to ensure that English instruction in aviation education supports not only certification requirements but also the communicative competence required for safe and effective operations.

Finally, this study reinforces the view that communicative authenticity is not merely a pedagogical concern but a core safety imperative in aviation education. Aligning TOEIC-based instruction with operational communication demands is therefore critical for enhancing instructional validity, strengthening pilot–ATC interaction, and contributing to the advancement of global aviation safety.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the Indonesian Flight Academy (API) Banyuwangi and the Ministry of Transportation of the Republic of Indonesia for granting research access and institutional support. Sincere thanks are extended to the English instructors and student participants whose collaboration made this study possible. This research was approved by the Research Ethics Committee of STKIP Al Hikmah Surabaya (Ref: 045/KEPK/STKIP-AH/IX/2024). The authors received no external funding, and all views expressed are solely their own.

## REFERENCES

Abeyratne, R. (2008). Language proficiency in aeronautical communications: Some recent developments. *Air and Space Law*, 33(4–5), 356–367. <https://doi.org/10.54648/aila2008027>

Alderson, J. C. (2009). Air safety, language assessment policy, and policy implementation: The case of aviation English. *Annual Review of Applied Linguistics*, 29, 168–187. <https://doi.org/10.1017/S0267190509090138>

Alderson, J. C. (2010). A survey of aviation English tests. *Language Testing*, 27(1), 51–72. <https://doi.org/10.1177/0265532209347196>

Aranda, W., & Henneberry, D. (2025). The role of simulated air traffic control environment (SATCE) in the implementation of ICAO level 4 requirements. *Advances in Human Factors and Ergonomics*, 186. <https://doi.org/10.54941/ahfe1006502>

Arslan, S. (2021). Examination of a computer-delivered English proficiency test for flight attendants: Test-takers' experiences. *Journal of Educational Technology and Online Learning*, 4(4), 673–687. <https://doi.org/10.31681/jetol.1003051>

Bystrova, B., Nemliy, L., Paziura, N., & Vasiukovych, O. (2019). Problem-based ESP methods for teaching future air traffic controllers to conduct radio exchange in non-routine situations. *Advanced Education*, 6(12), 74–79. <https://doi.org/10.20535/2410-8286.155041>

Clark, L., & Williams, G. (2020). English language proficiency in radiotelephony: A survey about its effect on the safety and efficiency of aviation. *The Especialist*, 41(4). <https://doi.org/10.23925/2318-7115.2020v41i4a9>

Coertze, S., Conradie, S., Burger, C., & Huddlestone, K. (2014). Aviation English in South African airspace. *Stellenbosch Papers in Linguistics Plus*, 42, 41–60. <https://doi.org/10.5842/42-0-167>

Demirdöken, G. (2024). Artificial intelligence in aviation English testing. *The Literacy Trek*, 10(3), 362–384. <https://doi.org/10.47216/literacytrek.1556603>

Dewi, T., & Ohi, S. (2023). Investigating English for aviation: An in-depth analysis of aviation English textbooks. *Journal of Teaching English for Specific and Academic Purposes*, 087. <https://doi.org/10.22190/JTESAP230301008D>

Farris, C., Trofimovich, P., Segalowitz, N., & Gatbonton, E. (2008). Air traffic communication in a second language: Implications of cognitive factors for training and assessment. *TESOL Quarterly*, 42(3), 397–410. <https://doi.org/10.1002/j.1545-7249.2008.tb00138.x>

Fatmawati, F., Rochmawati, L., & S, M. (2023). TOEIC-readiness and TOEIC mock tests: Measuring Civil Aviation Polytechnic of Surabaya cadets' preparedness for TOEIC test. *Jurnal Penelitian*, 7(4), 312–322. <https://doi.org/10.46491/jp.v7i4.1187>

Fauzi, I. (2020). EFL students' perception on TOEIC practice class at twelfth graders of SMK Kesehatan Husada Pratama. *Globish: An English-Indonesian Journal for English Education and Culture*, 9(2), 69–78. <https://doi.org/10.31000/globish.v9i2.2332>

Garcia, A., & Fox, J. (2020). Contexts and constructs: Implications for the testing of listening in pilots' communication with air traffic controllers. *The Especialist*, 41(4). <https://doi.org/10.23925/2318-7115.2020v41i4a4>

García, Á. (2023). Investigating the construct of aeronautical English listening testing: A qualitative analysis of the ICAO rating scale. *Journal of Teaching English for Specific and Academic Purposes*, 069. <https://doi.org/10.22190/JTESAP230220007G>

Guedes, D., & Coneglian, F. (2019). Tipos de questões para testes de proficiência das habilidades de leitura e escrita de língua inglesa. *Interfaces Científicas – Educação*, 7(3), 55–66. <https://doi.org/10.17564/2316-3828.2019v7n3p55-66>

Hautasaari, A., & Yamashita, N. (2015). Emotion detection in non-native English speakers' text-only messages by native and non-native speakers. In *Human-computer interaction* (pp. 183–200). Springer. [https://doi.org/10.1007/978-3-319-22701-6\\_13](https://doi.org/10.1007/978-3-319-22701-6_13)

Huang, H. (2013). Motivational changes in an English foreign language online reading context. *Social Behavior and Personality: An International Journal*, 41(5), 715–720. <https://doi.org/10.2224/sbp.2013.41.5.715>

Kanno, M. (2020). Apathy, agency, and sociocultural mediation in a university English course. *JALT PCP*, 144–156. <https://doi.org/10.37546/jaltpcp2019-18>

Kim, H. (2018). What constitutes professional communication in aviation: Is language proficiency enough for testing purposes? *Language Testing*, 35(3), 403–426. <https://doi.org/10.1177/0265532218758127>

Kim, H., & Elder, C. (2014). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing*, 32(2), 129–149. <https://doi.org/10.1177/0265532214544394>

Kraśnicka, I. (2016). English with flying colors: Aviation English and the International Civil Aviation Organization. *Studies in Logic, Grammar and Rhetoric*, 45(1), 111–124. <https://doi.org/10.1515/slgr-2016-0019>

Masrul, M., & Rasyidah, U. (2023). Washback effect of TOEIC listening and reading as a college exit test in Riau, Indonesia. *Journey (Journal of English Language and Pedagogy)*, 6(3), 698–712. <https://doi.org/10.33503/journey.v6i3.3767>

Masiulionienė, G., & Tupčiauskaitė, G. (2023). Bridging the gap between ICAO level 4 and 5 in aviation English: Expectations and results. *Aviation*, 27(3), 197–211. <https://doi.org/10.3846/aviation.2023.19763>

Mathews, E., Carson, J., Singleton, S., Valdes, E., & Williams, E. (2022). Investigating language factors in aviation accidents. *Aviation Psychology and Applied Human Factors*, 12(2), 99–108. <https://doi.org/10.1027/2192-0923/a000229>

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.

Oleksandrivna, K., & Kharlamova, L. (2019). Implementation of interactive teaching methods in the process of developing professional-communicative proficiency of future pilots. *International Journal of Science Annals*, 2(1–2), 12–18. <https://doi.org/10.26697/ijsa.2019.1-2.02>

Park, M. (2018). Innovative assessment of aviation English in a virtual world: Windows into cognitive and metacognitive strategies. *ReCALL*, 30(2), 196–213. <https://doi.org/10.1017/S0958344017000362>

Read, J., & Knoch, U. (2009). Clearing the air. *Australian Review of Applied Linguistics*, 32(3), 21.1–21.11. <https://doi.org/10.2104/ral0921>

Saputro, I., Nurhidayati, A., Suharno, S., Winiawati, C., Yunika, R., & Kusumadevi, M. (2023). Evaluation of the factory implementation of the CIPP method in Indonesian vocational education. *Al-Ishlah: Jurnal Pendidikan*, 15(4), 4972–4982. <https://doi.org/10.35445/alishlah.v15i4.3903>

Setyoko, A., Tunas, B., & Sunaryo, W. (2016). Evaluation of school operational assistance using the CIPP model in Indonesian private Islamic elementary schools. *International Journal of Managerial Studies and Research*, 4(3). <https://doi.org/10.20431/2349-0349.0403007>

Susanto, F., & Hamzah, S. (2022). The context, input, process, product (CIPP) evaluation model as a comprehensive framework for evaluating online English learning towards the industrial revolution era 5.0. *International Journal of Educational Research & Social Sciences*, 3(6), 2235–2247. <https://doi.org/10.51601/ijersc.v3i6.544>

Torres, J. C. S. (2023). Considerations for aviation-related English teachers' capacitation: A reflection about training aviation-specialized English instructors. *Journal of Teaching English for Specific and Academic Purposes*, 11(1), 57–67. <https://doi.org/10.22190/JTESAP230230006S>

Ulfah, B., & Mukminati, N. (2023). Bridging language proficiency and scientific knowledge: Integrating government science content into TOEIC-based summative listening tests. *Journal of English Language Studies*, 8(2), 339–353. <https://doi.org/10.30870/jels.v8i2.20014>

Vitryak, A., Slipak, B., & Serhii, K. (2016). Plain aviation English: Qualitative and quantitative interpretation. *Journal of Social Science Studies*, 4(1), 42–54. <https://doi.org/10.5296/jsss.v4i1.9478>

Wilson, K. (1989). Enhancing the interpretation of a norm-referenced second-language test through criterion-referencing: A research assessment of experience in the TOEIC testing context. *ETS Research Report Series*, 1989(2). <https://doi.org/10.1002/j.2330-8516.1989.tb00153.x>

Yin, Y., Yu, G., & He, D. (2023). Implementing the ICAO language proficiency requirements: The localization of aviation English testing. *Applied Linguistics*, 45(4), 699–717.  
<https://doi.org/10.1093/applin/amad057>

Zhao, K., Guo, X., & Gao, X. (2017). Learning English to fly: A study of Chinese cargo airline pilots' learning engagement. *English Today*, 33(4), 5–11.  
<https://doi.org/10.1017/S0266078417000219>