

## Bridging Language Proficiency and Scientific Knowledge: Integrating Government Science Content into TOEIC-Based Summative Listening Tests

Badriyah Ulfah <sup>a\*</sup>, Nur Mukminatien <sup>b</sup>

<sup>a</sup> Universitas Indo Global Mandiri, Palembang, Indonesia

<sup>b</sup> Universitas Negeri Malang, Malang, Indonesia

---

### Article Info

#### Article history

Submission Date: 25 May 2023

Acceptance Date: 11 September 2023

#### Keywords:

Listening Test; English for Specific Purpose; English for Government Science; TOEIC.

#### \*Corresponding author:

badriyahulfah@uigm.ac.id

### Abstract

To meet the demand for effective content integration in summative listening tests for English Specific Purpose classes catering to Government Science majors, this development project aims to create an engaging and immersive listening test. Drawing on the TOEIC-based framework, the test incorporates government science content as the thematic foundation of the questions. The development process comprises three key stages: exploration, development, and preliminary field testing. The test product was rigorously validated through expert assessment and a try-out test involving 50 students was conducted for preliminary field testing. The try-out test demonstrated a high reliability coefficient of 0.828. It encompasses 50 listening questions, spanning four sections: Photograph (10 questions), Questions-responses (16 questions), Conversation (12 questions), and Talks (12 questions). By seamlessly blending language proficiency with scientific knowledge, this innovative approach promises to enhance learning outcomes and bridge the gap between language acquisition and subject-specific competence for Government Science students. Based on the positive results obtained from the preliminary field testing and the validation process, it is recommended that the integrated TOEIC-based listening test be implemented in English Specific Purpose classes for Government Science majors.

© 2023 JELS and the Authors - Published by JELS.

---

## INTRODUCTION

The capacity to communicate in English has become an essential talent for everyone seeking job in today's globalized culture, particularly in global businesses or organizations. Many businesses and organizations use the Test of English for International Communication (TOEIC) to assess and evaluate their employees'

English proficiency (Doan & Hamid, 2021; Im, 2021; Park 2011). This test evaluates the listening and reading comprehension abilities of non-native English speakers in the workplace context. TOEIC scores itself provide a clear and objective measurement of a person's language skills, which can help employers make informed decisions about hiring and assigning job responsibilities. For instance, a high TOEIC score may suggest that a candidate has the language abilities needed to communicate with international clients or co-workers and is therefore qualified for a position that needs frequent English communication. A low TOEIC score, on the other hand, may indicate that an individual need additional language instruction or is unsuitable for specific work tasks, including as producing reports, presenting to an international audience, or attending international conferences (Oliveri & Tannenbaum, 2017). However, there are debates regarding the effectiveness and limitations of TOEIC as a tool for measuring English proficiency.

TOEIC has limitations when assessing an individual's English language proficiency, despite its benefits. The TOEIC is regarded as a test of a situation's general theme, regardless of the test-takers prior content knowledge. Some scholars (Powers & Powers, 2015; Sirikanjanawong & Wasanasomsithi, 2018) argue that TOEIC does not accurately reflect real-life workplace communication situations for instance, the test may not assess a person's ability to understand accents or dialects that differ from those commonly found in TOEIC materials. Another limitation of TOEIC is that it may not consider the cultural and contextual differences in workplace communication. For example, a person who has grown up speaking English may have an advantage in understanding certain cultural references or idioms that are commonly used in the workplace they are applying to, while someone who has learned English as a second language may struggle with these nuances (Kim & Haneda, 2020). In short, TOEIC is a widely recognized and standardized test that can provide valuable information about an individual's English language proficiency in the global workforce. Hence, it is important to recognize its limitations and use it in conjunction with other methods for a more comprehensive evaluation of an individual's language abilities.

One potential improvement for TOEIC could be the addition of topic-specific tests. This would allow individuals to demonstrate their language proficiency in a specific field, such as government science, and provide employers with more targeted information about an individual's language abilities in relation to their job

responsibilities. Additionally, this could help to address some of the limitations of TOEIC by assessing an individual's language proficiency in more relevant workplace communication situations (Daza & Suzuki, 2004; Schmidgall & Powers, 2020). Therefore, Further research and development are needed to improve the accuracy and relevance of language proficiency assessments in the global workforce (de Jong, 2023; Knoch & Macqueen, 2016). This can help to ensure that individuals are being evaluated on their language abilities in a way that is fair and relevant to their specific job responsibilities. To achieve this, collaboration between language experts and professionals in different fields is necessary to create topic-specific tests that accurately assess an individual's language abilities in specific fields or job responsibilities. Overall, it is important to remember that while standardized language tests like TOEIC can be useful, they are just one piece of the puzzle when evaluating an individual's language abilities in the global workforce. Therefore, the research question that this study seek is to answer how TOEIC-based summative listening test is developed by integrating content subject, specifically government science, for English for International Communication (EIC) 4 class at a private University in Indonesia. The result of the study would then present in the form of a listening test set where subject content namely government science is integrated

### **Test of English for International Communication (TOEIC)**

The Educational Testing Service (ETS) introduced the TOEIC exam in 1979, and it has since become one of the most extensively utilized English proficiency examinations in the world. The objective of the exam is to evaluate the test-ability takers to communicate in English in an international business setting (ETS, 2013, 2015a; 2016a). The test is utilized by multinational enterprises, educational establishments, and governmental organizations to assess the level of English language aptitude among individuals who are not native speakers of English. (ETS, 2015b, 2016b). The Test of English for International Communication (TOEIC) is a standardized assessment that aims to assess the English language proficiency of individuals who are not native speakers in a global professional context. The test has a total score range of 10 to 990 and consists of two sections: listening and reading. The listening segment consists of four sections and a variety of listening exercises, including picture, short conversations, talks, and question-and-answer sessions. In 45 minutes, the test-taker must respond to 100 multiple-choice questions. The reading segment consists of three sections and a variety of reading exercises, including brief

passages, business letters, and reports. In 75 minutes, the examinee must respond to 100 multiple-choice questions. The test scores range from 10 to 990, and they are valid for two years (ETS,2015b; ETS 2016b). The listening component evaluates the test-ability takers to comprehend English spoken in a variety of accents and contexts, and the reading section evaluates their comprehension of written English (Im & Cheng, 2019). While being widely recognized and utilized by international corporations and organizations, the TOEIC has been criticized for failing to evaluate the test-topic taker's knowledge, which may result in an erroneous assessment of their English language competency.

One of the main criticisms of the TOEIC is that it does not assess the test-subject-matter taker's expertise. For instance, a test-taker who is a subject matter expert in a certain sector yet has a strong command of English in their own domain may not perform well on the test. The test is designed to evaluate overall business English competence (ETS, 2017), but it may not fully reflect the test-ability takers to communicate in a specific industry or field. Another critique of the TOEIC is that it may not fully represent the linguistic diversity of English spoken around the world (Schmidgall & Powers, 2020). The test focuses mostly on business-related English, which may not accurately reflect the language used in other contexts, such as social interactions or academic settings. In conclusion, the TOEIC test is an essential instrument for assessing the English language ability of non-native English speakers in a corporate setting. However, it does not take into consideration the test-topic taker's knowledge and does not adequately reflect the linguistic diversity of English spoken around the world. Notwithstanding these complaints, the TOEIC test continues to be widely known and utilized.

### **English for Specific Purpose of Government Science**

English for Specific Purposes (ESP) is an aspect of English language acquisition that emphasizes imparting English language skills pertinent to a particular field or profession. It is tailored to satisfy the needs of students in particular fields or disciplines, such as engineering, medicine, law, and government science, etc (Hutchinson & Waters, 1989). The teaching of language skills in ESP is contextualized, prioritizing the development of communicative competence and the effective utilization of English within specific situations. (Bui, 2022; Gaffas, 2019; Huang & Yu, 2023). In the field of government science, ESP is particularly important as it involves the use of English in a wide range of contexts, such as policy-making, diplomacy,

international relations, and public administration (Kareva, 2013). The use of English in these contexts requires a specialized vocabulary, discourse patterns, and communication strategies, which are not always taught in general English language courses. Therefore, the objective of ESP for government science is to equip students with the language skills necessary for effective communication in their field of employment.

In the context of government science, listening skills are particularly important as learners are often required to understand and interpret complex policy discussions, negotiations, and presentations. In ESP for government science, therefore, the development of listening skills is a crucial aspect of language learning. Listening materials are frequently drawn from authentic sources, such as policy addresses, parliamentary debates, and news broadcasts, in order to expose learners to the types of listening situations they may encounter in their respective fields (Djelloul & Melouk, 2022). However, the TOEIC test does not assess learners' content knowledge or their ability to understand specialized vocabulary related to their field of study. Furthermore, the development of ESP for government science is influenced by various factors such as the learners' educational background (Poedjiastutie, 2017), job requirements (Tong & Gao, 2022), and the specific language needs of their field of work (Bunch & Martin, 2021). Therefore, the design of ESP programs for government science should be tailored to the specific needs and goals of the learners and their field of work (Iswati & Triastuti, 2021; Ulfah et al, 2021). This includes the selection of appropriate materials, tasks, and assessments that are relevant and authentic to their field of study (Rao, 2019). Consequently, the integration of content subject into a TOEIC-based summative listening test can also provide teachers and curriculum designers with valuable feedback for improving and refining ESP programs to meet the evolving requirements and challenges of learners in the field of government science (Lee, 2020).

## **METHOD**

### **Participants**

English proficiency is necessary for the students to compete on the global job market, and this proficiency is frequently verified by the Test such as the Test of English for International Communication (TOEIC). To prepare students for this test, many higher educations have integrated it into their curriculum and adopted it as a

summative assessment for their English classes. In Indonesia, English has considered to necessary subject for higher education students from all major. One of the private universities in South Sumatera, Indonesia which offers English for Specific Purpose (ESP) classes to all its students. These classes are intended to provide students with the English language skills necessary for success in their chosen disciplines, while also preparing them for the TOEIC examination. Specifically, students are required to take English for Specific class named English for International Communication (EIC) 1, 2, 3, and 4, which are designed to improve their communication skills in English for the international workplace. Therefore, in occupying the needs of summative test to EIC class which also incorporating the TOEIC test, it is needed to developing the summative test that integrating the content subject of the students into the TOEIC-based test for these classes.

This study focused on the development of a TOEIC-based summative listening test for EIC 4 students in the university. The test is developed by integrating the content subject of the students, which in this case is government science, into the test questions. The development process follows the steps outlined by Borg and Gall (1987) for development research. The test followed the TOEIC structure test which consists of 50 listening questions divided into four sections namely Photograph, Questions-responses, Conversation, and Talks. Each question is designed to measure the students' English language proficiency as well as their knowledge of government science. Therefore, the test takers are the students of Government Science study program in the university who had passed EIC 1, EIC 2, and EIC 3 class and have finished the 15 meeting of EIC 4 classes. A total of 50 students who were currently taking EIC 4 class would join the try-out test for the preliminary field testing.

### **Data Collection & Analysis**

In developing the Summative Listening test, the writer use Borg and Gall (1987) Research & Development model as the reference. In this model, there are 10 steps promoted for developing a test namely Research and collection preliminary, Research Planning, Early Product Development, Expert Validation, Product Revision, Early Test, Product Revision, Field Test, Final Product Revision and Dissemination. But for this listening test development procedure, the researchers would follow the model with some adjustment, so the procedures that were done as below:

#### *1. Pre-Development (Research and Information Collecting)*

In this step, the researchers would first examine English for International Communication (IEC) course description and formulate the indicators for the test. Because the test would also be adjusted to the requirements of TOEIC, the researchers would also prepare the comparison of the test in the form of table between the test items that the researchers would develop and the TOEIC.

2. *Developing (realization of the product)*

In this step, the researchers would start to make the test items by referring to the results of the first step that have been done.

3. *Producing the Final Product*

After finishing the second steps and the 50 test items are ready, I did the validity and reliability testing. The content validity employed two expert judgments. The researchers employed the experts (English lecturers) to review the test items and rate their relevance to the construct being measured. The experts can provide feedback on whether the test items cover all aspects of the construct and whether they are appropriate for the intended population of Government Science students. And then, the 50 test items would be given to the students in order to measure the Reliability. The researchers would use Internal consistency reliability which measures the consistency of scores across different items within a test. The researchers would use statistical techniques of Cronbach alpha and split-half reliability with the help of SPSS program for the analysis

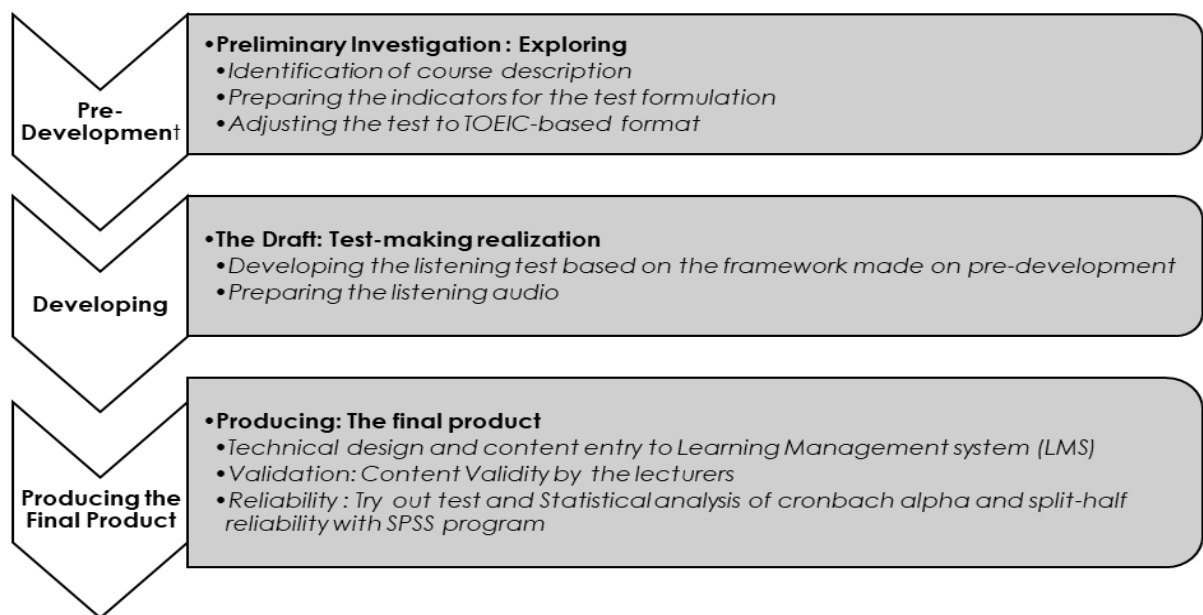


Fig 1. Procedure of Test Development

## RESULT AND DISCUSSION

As it has been mentioned on the methodology, the researchers did the preliminary investigation the EIC 4 course in order to match the needs of the online listening test. The result of the class description is presented on table 1.

**Table 1. Class Description**

Course Description		The EIC 4 course is an advanced level course that builds upon the foundational knowledge and skills acquired in the preceding EIC 1, 2, and 3 courses, with a focus on English language proficiency for international communication. The primary objective of this course is to enhance and fortify the English language proficiency of students in the field of government science. The course aims to equip students with the ability to effectively utilize listening, speaking, reading, and writing skills in English, while adhering to appropriate grammar and vocabulary standards. The EIC 4 course serves as the ultimate stage of readiness for students in anticipation of their TOEIC English proficiency assessment.
Content Subject Learning Objectives	1	Mastering the basic concepts and theories of government, political theory, bureaucracy, state, democracy, and the public sector.
	2	Mastering the theoretical concepts of political institutional systems, elections, and parliament
	3	Mastering the theoretical concepts of managing government human resources, managing state/public finances and state assets.
	4	Mastering the theoretical concepts of society and the diversity of Indonesian society and culture
Course Objectives	1	Students are able to listen, comprehend and re-tell the description of a picture that occurs in governmental science situation by using good and correct English vocabulary and grammar
	2	Students are able to listen, comprehend, and give response to a simple conversation about government science
	3	Students are able to listen, comprehend, and give response to a simple discussion about government science
	4	Students are able to analyze the grammatical and vocabulary in reading texts on the topic of government science
	5	Students are able to write a simple essay on the topic of governance by using correct grammatical and vocabulary
Teaching Method		Online Learning
Assessment Method		Online Assessment with Learning Management System (LMS)

Based on the information provided on table 1, the researchers do the next step of Preparing the indicators for the test formulation and adjusting the test to TOEIC-based format. The result is presented on table 2.



**Table 2. Test Specification**

Test Taker	The students of Government Science study program who passed IEC 1, IEC 2, and IEC 3 classes and have finished the 15 meetings in IEC 4 classes.		
Test Objective	<ol style="list-style-type: none"> <li>1. To measure the students' listening ability in understanding the descriptive utterance</li> <li>2. To measure the students' listening ability in giving response to simple question about government science.</li> <li>3. To measure the students' listening ability in understanding details in short spoken exchanges</li> <li>4. To measure the students' listening ability in understanding the central idea, purpose, and basic context of extended spoken texts</li> </ol>		
Time Allocation	100 minutes		
Item Formats	Multiple Choice		
Structure of the test	Photograph		
*TOEIC-based format (ETS, 2017)	Questions & Response Conversation Talks		
Number of Questions	50 questions		
Scoring	Correct answer x 2		
Grading (*Based on the university regulation)	>= 80 s/d 100	Grade A	Scale 4
	>=68 s/d 79	Grade B	Scale 3
	>= 56 s/d 67	Grade C	Scale 2
	>= 46 s/d 55	Grade D	Scale 1
	0 s/d 45	Grade E	Scale 0

**Proposed Listening Test Theme**

The product of this research is a set of listening test consists of 50 multiple questions. Since this EIC class is projected to be TOEIC - based test, there were four sections of the test following TOEIC proficiency test namely photograph, question & response, conversation, and talks (ETS, 2017). First, in Part 1, there are 10 questions about Photograph, in each question, the students looked at a photograph and then listen to 4 sentences. The students should choose the sentence that best describes the photograph. Second, in Part 2, there are 16 questions about Questions & Responses, in each question, the students listened to a question and then listen to 3 possible responses. The students choose the correct response. Third, in part 4, there are 12 questions about Conversation, in this part, the students listened to three short conversations, each with 4 questions. For each question, the students choose the answer which fits best to the conversations. Lastly, in Part 4, there are 12 questions about talks, in this part, the students are going to listen to four short talks, each with 3 questions. For each question, the students choose the answer which fits best to the talks. The detail information is provided in table 3.

**Table 3. Test Structure**

Test Items	Direction	Objective of the test	Questions Number	Total Questions
Part 1: Photograph	For each question in this part, you will hear four statements about a picture in your test book. When you hear the statements, you must select the one statement that best describes what you see in the picture. Then find the number of the question on your answer sheet and mark your answer. The statements will not be printed in your test book and will be spoken only one time.	To measure the students' listening ability in understanding the descriptive utterance	1 - 10	10
Part 2: Question & Response	You will hear a question or statement and three responses spoken in English. They will not be printed in your test book and will be spoken only one time. Select the best response to the question or statement and mark the letter (A), (B), or (C) on your answer sheet	To measure the students' listening ability in giving response to simple question about government science	11 – 26	16
Part 3: Conversation	Directions: You will hear some conversations between two or more people. You will be asked to answer four questions about what the speakers say in each conversation. Select the best response to each question and mark the letter (A), (B), (C), or (D) on your answer sheet. The conversations will not be printed in your test book and will be spoken only one time.	To measure the students' listening ability in understanding details in short spoken exchanges	27 – 38	12
Part 4: Talks	You will hear some talks given by a single speaker. You will be asked to answer four questions about what the speaker says in each talk. Select the best response to each question and mark the letter (A), (B), (C), or (D) on your answer sheet. The talks	To measure the students' listening ability in understanding the central idea, purpose, and basic context of extended	39 – 50	12

---

will not be printed in your test book and will be spoken only one time.

---

As it has been mentioned, this developed listening test would integrate the content knowledge of the government science to the English Summative Listening Test for the EIC class. The distribution of the content subject to the listening questions is described in table 4.

**Table 4. Content subject Integration Distribution**

<b>Content Subject</b>	<b>Total Number of Questions</b>	<b>Distribution</b>	<b>Question Number</b>
Government Bureaucracy	19	Part 1 (Photograph): 6 Questions	Q1, Q2, Q5, Q7, Q8, Q10
		Part 2 (Question & Response): 5 Questions	Q12, Q13, Q17, Q21, Q24
		Part 3 (Conversation): 4 Questions	Q 31, Q32, Q33, Q34
		Part 4 (Talks): 4 Questions	Q39, Q40, Q41, Q42
Public Leadership	6	Part 1 (Photograph): 1 Question	Q9
		Part 2 (Question & Response): 1 Question	Q20
		Part 4 (Talks): 4 Questions	Q43, Q44, Q45, Q46
Democracy	7	Part 1 (Photograph): 1 Question	Q6
		Part 2 (Question & Response): 2 Questions	Q19, Q26
		Part 3 (Conversation): 4 Questions	Q27, Q28, Q29, Q30
Politik Party	3	Part 1 (Photograph): 1 Question	Q4
		Part 2 (Question & Response): 2 Questions	Q14, Q23
Public Economy	6	Part 1 (Photograph): 1 Question	Q3
		Part 2 (Question & Response): 5 Questions	Q15, Q16, Q18, Q22, Q25
Society and Culture	9	Part 1 (Photograph): 1 Question	Q11
		Part 3 (Conversation): 4 Questions	Q35, Q36, Q37, Q38
		Part 4 (Talks): 4 Questions	Q47, Q48, Q49, Q50

**Preliminary Field-Testing Result  
Validity**

Content validation is a rigorous process that involves experts in the field evaluating the test items to ensure that they align with the intended construct being measured. If the results of the summative listening test are deemed valid through content validation, this indicates that the test accurately measures what it is intended to measure. In this case, the two English lecturers serving as validators bring their expertise in the English language and listening comprehension to the table to assess the test items for accuracy and relevance. If the test passes content validation, it suggests that the items are appropriately designed and phrased to measure the listening skills of the test takers. This can provide assurance to stakeholders such as educators, administrators, and policymakers that the results of the test are reliable and meaningful. Furthermore, the fact that two expert validators were involved in the process lends credibility to the results and supports the notion that the test is a valid measure of listening comprehension. Overall, passing content validation provides a strong indication that the summative listening test is an effective tool for assessing listening skills.

The assessment of content validity for our summative listening test for Government Science majors was conducted meticulously using a checklist adapted from Mora (2020). This checklist comprised six key dimensions: content (7 questions), reliability (8 questions), authenticity (4 questions), appropriateness (4 questions), impacts (5 questions), and practicality (4 questions). The two experienced validators independently evaluated each dimension by providing "Yes" or "No" responses to assess the alignment of the test with the specified criteria. It is important to note from the result said that some checklist items, such as those related to reliability, may not receive responses from validators when certain conditions, like video availability, were not applicable. It's noteworthy that, in response to constructive feedback, it is needed to add example questions for clarification in each section of the test. This addition aims to enhance the comprehensibility and user-friendliness of the assessment, ensuring that test-takers and instructors have a clear understanding of the expected format and content. Furthermore, it should be noted that some checklist items, particularly those related to reliability, might not receive responses from validators when specific conditions, such as video availability, are not applicable

The Content Validity Ratio (CVR) formula was then applied to each dimension, where  $CVR = \frac{\text{Number of "Yes" Responses} - (\text{Number of Validators}/2)}{\text{Number of Items}}$

(Number of Validators/2). This formula helped the researchers quantify the degree of consensus among validators regarding the importance of each dimension for test validity (Almanasreh et al, 2019). CVR values above 0.5 indicated substantial agreement among validators. The content validity assessment yielded promising results, reaffirming the robustness of our summative listening test. Across all dimensions, the calculated CVR values indicated strong consensus among validators, with most dimensions achieving CVR values well above the 0.5 threshold. Notably, the dimensions of validity, reliability, authenticity, appropriateness, and impacts, and practicality all garnered CVR values of 6, 5, 3, 3, 4, and 3, respectively. These high CVR values demonstrate a clear consensus among validators that the test effectively measures language proficiency and subject-specific competence. While practicality received a CVR of 3, it still demonstrated agreement among validators regarding the feasibility of test construction and administration. In summary, the content validity results affirm the test's ability to accurately assess the targeted skills and knowledge of Government Science majors, reinforcing its utility within English for Specific Purposes (ESP) classes tailored to this specialized academic context.

### **Reliability**

The internal consistency reliability refers to the consistency of scores across different items within a test. In this case, the researchers used statistical techniques of Cronbach's alpha, split-half reliability, and other measures to assess the reliability of the test. The Cronbach's alpha coefficient is a frequently employed to measure the internal consistency reliability that indicates the degree to which a group of items is related. In this study, the Cronbach's alpha coefficient for questions 1 through 25 was 0.805 and for questions 26 through 50 it was 0.814. These values indicate a high level of internal consistency among the items, as scores across the different items within each part of the test are strongly related. The correlation coefficient between the two forms of the test was 0.708, indicating a moderate positive relationship between the two forms. This suggests that the two forms of the test are measuring similar constructs, but there may be some differences between the two versions. The Spearman-Brown coefficient, a measure of the test's reliability when prolonged or shortened, was 0.829. This indicates that the test's reliability remained consistent regardless of the duration of the examination. The Guttman split-half coefficient, which evaluates the reliability of a test by dividing it in half, was also 0.829. This indicates that the test is reliable even when it is split into two halves. Overall, these

results suggest that the summative listening test has a high level of internal consistency and is a reliable measure of listening comprehension. The researchers can be confident that the test scores are consistent and accurate and can use the test to make valid inferences about the listening comprehension abilities of test-takers. The detail result is presented on table 5.

Table 5. Reliability test Statistics

Cronbach's Alpha	Part 1	Value	.805
		N of Items	25 <sup>a</sup>
	Part 2	Value	.814
		N of Items	25 <sup>b</sup>
	Total N of Items		50
Correlation Between Forms			.708
Spearman-Brown Coefficient	Equal Length		.829
	Unequal Length		.829
Guttman Split-Half Coefficient			.828

## CONCLUSION

In conclusion, this development project addresses the need for content subject integration in summative listening tests designed for English Specific Purpose classes catering to Government Science majors. The resulting online listening test, based on the TOEIC framework and enriched with government science content, has undergone a meticulous development process encompassing exploration, development, and preliminary field testing for validity and reliability. Comprising 50 questions spanning four sections, namely Photograph, Questions-responses, Conversation, and Talks, the test has been validated by English lecturers, attaining high levels of validity and reliability.

The implications of this project are significant for English language educators instructing Government Science major students. Integrating content subject material into listening tests offers students a more authentic language learning experience, preparing them for real-world situations they may encounter in their future careers. Furthermore, the use of government science content enhances students' subject knowledge, crucial for academic and professional success. The online listening test also provides a practical means for assessing listening comprehension abilities efficiently.

This study underscores the importance of content validation and reliability testing in the development of summative listening tests to ensure their validity,

reliability, and accuracy. Based on the findings, the researchers recommend that English language instructors in ESP courses for Government Science majors consider adopting content integration in their listening tests. The developed TOEIC-based online listening test, incorporating government science content, can serve as a model for similar assessments in various subject areas. Future research endeavors may explore the impact of such tests on students' listening comprehension skills and their academic and professional achievements. Additionally, investigating alternative measures of validity and reliability in summative listening test creation could further enhance assessment practices in the field.

## REFERENCES

- Almanasreh, E., Moles, Rebekah.,M., & Chen, T.F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Bui, H.P. (2022). Students' and teachers' perceptions of effective ESP teaching. *Heliyon*, 1-7. <https://doi.org/10.1016/j.heliyon.2022.e10628>
- Bunch, G.C & Martin, D. (2021). From "academic language" to "language of ideas": A disciplinary perspective on using language in K-12 settings. *Language and Education*, 32(6), 539-556. <https://doi.org/10.1080/09500782.2020.1842443>
- Daza, C., & Suzuki, M. (2004). A review of the reading section of the TOEIC. *TESL Canada Journal*, 22(1), 16–24. <https://doi.org/10.18806/tesl.v22i1.163>
- de Jong, N.H. (2023). Aseing second language speaking proficiency. *Annual Review of Linguistics*, 9, 541-560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>
- Djellol, K.H. & Melouk, M. (2022). Meeting students' needs: ESP teaching at the department of political sciences. *African Education Research Journal*, 10 (2), 150 – 160. <https://doi.org/10.30918/AERJ.102.22.001>
- Doan, L.D. & Hamid, M. O. (2021). Economics of English: Examining the demand for English proficiency in the Vietnamese job market. *RELC Journal*, 52(3), 618-634.
- Educational Testing Service. (2013). *User guide – Listening and reading*. Retrieved from [www.ets.org/Media/Tests/Test\\_of\\_English\\_for\\_International\\_Communication/TOEIC\\_User\\_Gd.pdf](http://www.ets.org/Media/Tests/Test_of_English_for_International_Communication/TOEIC_User_Gd.pdf).
- Educational Testing Service. (2015a). *Examinee handbook – Listening & reading*. Retrieved from [www.ets.org/Media/Tests/TOEIC/pdf/TOEIC\\_LR\\_examinee\\_handbook.pdf](http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_LR_examinee_handbook.pdf).
- Educational Testing Service. (2015b). *Sample questions of reading comprehension*. Retrieved from [http://exam.ybmnet.co.kr/toEIC/info/sample\\_rc5.asp](http://exam.ybmnet.co.kr/toEIC/info/sample_rc5.asp).
- Education Testing Service (2016a). *Examinee handbook– Listening & reading*. Retrieved from [www.etsglobal.org/content/download/15388/261514/version/5/file/TOEIC+LR+and+SW+Examinee+Handbooks.pdf](http://www.etsglobal.org/content/download/15388/261514/version/5/file/TOEIC+LR+and+SW+Examinee+Handbooks.pdf).

- Educational Testing Service. (2016b). *The TOEIC® tests – The global standard for assessing English proficiency for business*. Retrieved from [www.ets.org/toEIC/succeed](http://www.ets.org/toEIC/succeed).
- Educational Testing Service. (2017). *Examinee handbook – For the updated version of the TOEIC® listening and reading test*. Retrieved from [www.ets.org/s/toEIC/pdf/examinee-handbook-for-toEIC-listening-reading-test-updated.pdf](http://www.ets.org/s/toEIC/pdf/examinee-handbook-for-toEIC-listening-reading-test-updated.pdf).
- Gaffas, Z.M. (2019). Students' perceptions of the impact of EGP and ESP courses on their English language development: Voices from Saudi Arabia. *Journal of English for Academic Purposes*, 42. <https://doi.org/10.1016/j.jeap.2019.100797>
- Huang, Q. & Yu, Q.J. (2023). Towards a communication-focused ESP course for nursing students in building partnership with patients: A needs analysis. *English for Specific Purposes*, 70, 57-69. <https://doi.org/10.1016/j.esp.2022.11.006>.
- Hutchinson, T & Waters, A. (1989). *English for specific purposes: A learning-centred approach*. Cambridge: Cambridge University Press
- Im, G. (2021). What do international business professionals say about the TOEIC. *Journal of The Korea English Education Society*, 20(1), 1-37.
- Iswati, L & Triastuti, A. (2021). Voicing the challenges of ESP teaching: Lessons from ESP in Non-English departments. *Studies in English Language and Education*, 8(1), 276-293. <https://doi.org/10.24815/siele.v8i1.17301>
- Kareva, V. (2013). English for specific purposes: Public administration and political sciences. *Procedia Social and Behavioral Sciences*, 70, 1477 - 1481. <https://doi.org/10.1016/j.sbspro.2013.01.214>
- Kim, M. & Haneda, M. (2020). The disconnect between English tests and English proficiency: Two South Korean jobseekers' perceptions and performances. In Mirhosseini, S. & Costa, P.I, (Eds), *The socio politics of English language testing* (pp.147-164). London: Bloomsbury Publishing Plc
- Knoch, U. & Macqueen, S. (2016). Language assessment for the workplace. In Tzagari, D. & Banerjee, J. (Eds.), *Handbook of Second Language Assessment* (pp 291-308). De Gruyter Mouton
- Lee, J. S. (2020). The role of informal digital learning of English and a high-stakes English test on perceptions of English as an international language. *Australasian Journal of Education Technology*, 36(2), 155-168. <https://doi.org/10.14742/ajet.5319>
- Mora, J. (2020). Checklists for validating an English test in the pre-testing stage. The 11<sup>th</sup> Hatyai National and International Conference (pp. 2032-2041). Retrieved from [https://www.hu.ac.th/conference/proceedings2020/doc/G13/G13-4-\[1006Ed-OP\]%20\(P-2032-2041\).pdf](https://www.hu.ac.th/conference/proceedings2020/doc/G13/G13-4-[1006Ed-OP]%20(P-2032-2041).pdf)
- Oliveri, M.E & Tannenbaum, R.J. (2017). Insight into using TOEIC® test scores to inform human resource management decisions. *ETS Research Report Series*, 1(1).
- Park, J.S. (2011). The promise of English: Linguistic capital and the neoliberal worker in the South Korean job market. *International Journal of Bilingual Education and Bilingualism*, 14(11), 443-455.



- Poedjiastutie, D. (2017). The pedagogical challenges of English for specific purposes (ESP) teaching at the university of muhammadiyah malang, Indonesia. *Educational Research and Reviews*, 12(6), 338-349. <https://doi.org/10.5897/ERR2016.3125>
- Powers, D.E & Powers, A. (2015). The incremental contribution of TOEIC® listening, reading, and writing tests to predicting performance on real-life English language tasks. *Language testing*, 32(2), 151-167.
- Rao, P.S. (2019). The effective use of authentic materials in the English language classrooms. *International Journal of Arts, Science and Humanities*, 7(1), 1-8. <https://doi.org/10.34293/sijash.v7i1.556>
- Schmidgall, J. & Powers, D.E. (2020). TOEIC writing test scores as indicators of the functional adequacy of writing in the international workplace: Evaluation by linguistic laypersons. *Assessing Writing*, 46. <https://doi.org/10.1016/j.asw.2020.100492>
- Sirikanjanawong, N & Wasanasomsithi, P. (2018). Relationship between the ICAO language proficiency requirements (LPRs) and test of English for international communication (TOEIC) scores of flight attendants in Thailand. *LEARN Journal: Language Education and Acquisition Research Network*, 11(1), 64-86.
- Tong, M & Gao, T. (2022). For sustainable career development: Framework and assessment of the employability of business English graduates. *Frontiers in Psychology*, 13, 1-14. <https://doi.org/10.3389/fpsyg.2022.847247>
- Ulfah, B., Petrus, I., Mirizon, S. (2021). ESP Needs-based course design for the employees of government protocol department. *English Language in Focus*, 3(2), 143-158. <https://doi.org/10.24853/elif.3.2.143-158>

## APPENDICES

1. Listening Test Questions Sheet can be accessed on

[https://drive.google.com/file/d/1-l\\_BagyT40nKaSvMqQuVIRR7E96HDArE/view?usp=sharing](https://drive.google.com/file/d/1-l_BagyT40nKaSvMqQuVIRR7E96HDArE/view?usp=sharing)

2. Listening Audio Files and Script can be accessed on

<https://drive.google.com/file/d/1zdRKCYcuPVVA5SaM05h33bvxxgA6CkSO/view?usp=sharing>