

What Are the Effective Tests for Measuring English Proficiency in a Japanese Project-Based English Program? A Comparison of the Test of English for International Communication, Global Test of English Communication, and CNN Global English Test Service

Jun Sakaue¹ & Tsukasa Yamanaka²

¹ Language Education Center, Ritsumeikan University, Kyoto, Japan

² College of Life Sciences, Ritsumeikan University, Kusatsu, Japan

Correspondence: Tsukasa Yamanaka, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, 525-8577, Japan. E-mail: yaman@fc.ritsumei.ac.jp

Received: September 21, 2024

Accepted: October 22, 2024

Online Published: October 24, 2024

doi: 10.5539/elt.v17n11p81

URL: <https://doi.org/10.5539/elt.v17n11p81>

Abstract

Despite significant efforts to improve English education, challenges such as declining motivation and slow progress persist. In this study we examine the relationships among three English proficiency tests—Test of English for International Communication (TOEIC), Global Test of English Communication (GTEC), and CNN Global English Test Service (CNN GLENTS)—in the context of university-level English education in Japan. We aim to explore the effectiveness of these tests by comparing scores and improvements across various sections. The analysis reveals notable differences in the constructs measured by each test. CNN GLENTS exhibits weak or negative correlations with GTEC and TOEIC, suggesting divergent skill assessments, particularly in the international studies section. We observe high correlations between the listening and reading sections of GTEC and TOEIC in the second round, possibly due to increased test familiarity, whereas correlations involving speaking and writing are weaker. These findings indicate that CNN GLENTS emphasizes different proficiencies, such as background knowledge and contextual understanding, which GTEC and TOEIC do not directly capture. The study underscores the need for a multifaceted approach to language assessment, given no single test fully captures the complexity of English proficiency. Future researchers should explore the underlying factors that influence these correlations to improve the utility and interpretation of these assessment tools in educational settings.

Keywords: project-based English program, TOEIC, GTEC, CNN GLENTS, language proficiency assessment

1. Introduction

In this study we aim to explore more effective ways to utilize the Test of English for International Communication (TOEIC), Global Test of English Communication (GTEC), and CNN Global English Test Service (CNN GLENTS) by comparing and analyzing the scores and improvements across the sections of TOEIC, GTEC, and CNN GLENTS in university English education. Over the years, concerns have persisted regarding the effectiveness of English education in Japan (see Butler & Iino, 2005). University-level English education, the focus of this study, is no different. Despite significant governmental efforts, the goal of creating ‘English-speaking Japanese’ (Eigo ga Tsukaeru Nihonjin) (Ministry of Education, Culture, Sports, Science and Technology of Japan, 2002) remains unfulfilled, as challenges such as declining motivation post enrollment, slow progress in language proficiency, and low standings in international rankings continue to hinder progress.

To address these persistent challenges, various initiatives have been implemented, including project-based learning approaches like the project-based English program (PEP) promoted by the authors. In PEP, students select a topic based on their own interests and conduct experiments or research to make new discoveries, which they then present in the form of an academic presentation. The aim of the program is for students to develop their English language skills through engaging in these activities (Yamanaka et al., 2021).

Although PEP evaluates student progress using the TOEIC Listening & Reading (LR) test, the program’s emphasis on presentations highlights the need for assessments that measure skills beyond LR. Therefore, in this

study we aim to explore more effective evaluation methods by comparing and analyzing the scores and improvements across the sections of the GTEC, which assesses LR, writing, and speaking skills; the CNN GLENTS, which also evaluates background knowledge; and the TOEIC. Understanding the strengths and limitations of each testing method will provide insights into designing more effective language education programs, ultimately contributing to the overarching goal of creating ‘English-speaking Japanese.’

2. Literature Review

2.1 Three Tests Used in the Experiment

We compared three tests designed to measure English proficiency. We present a summary of each test as follows.

2.1.1 TOEIC

The TOEIC developed by Educational Testing Service (ETS), is widely recognized for its focus on practical English skills relevant to professional settings (ETS, 2024). The test is divided into two main sections: Listening and Reading, with a combined total score range of 10–990. Researchers have highlighted the TOEIC’s validity and reliability, emphasizing its role in assessing language skills necessary for workplace communication (Powers & Powers, 2015). The TOEIC test is available in two main formats: the public version and the Institutional Program (IP) version. The public version, administered at official test centers, provides a standardized measure of English proficiency, specifically designed to assess LR skills in an international context. In contrast, the TOEIC IP version is administered within institutions, such as universities and companies, and offers a more flexible, cost-effective option for internal assessments. In this study, students took the TOEIC IP version, which can be taken individually online.

2.1.2 GTEC

In contrast, the GTEC, developed by Benesse Corporation, provides a more comprehensive evaluation framework. It comprises four components: LR, writing, and speaking. The total score ranges from 0–1000, with each component contributing up to 250 points. This test aims to provide a comprehensive measure of English proficiency by assessing a range of language skills in both academic and everyday contexts (Benesse Corporation, n.d.). The GTEC assessment system is structured to cater to different educational stages, with ‘GTEC-Junior’ designed for elementary and junior high school students, ‘GTEC’ for junior high and high school students, ‘GTEC-Academic’ for university students and adults, and ‘GTEC-Business’ for professionals in the workplace. Because the participants in this study were university students, we selected GTEC-Academic. Researchers have explored GTEC’s effectiveness in various educational settings, noting its balanced assessment across different language skills (Kim et al., 2017).

2.1.3 CNN GLENTS

The CNN GLENTS, developed by Asahi Press, is a relatively recent entrant in the field of English language proficiency assessments. Launched in 2018, it integrates real-world news content and global issues from CNN’s media platform into its test materials, offering a unique approach that assesses how well candidates can understand and analyze contemporary topics in English. The listening section is scored out of a maximum of 60 points, and the reading section is scored out of a maximum of 40 points, making a total of 100 points. Additionally, there is an international studies section scored out of 20 points; however, this section is excluded from the assessment of ‘English proficiency.’ One key feature of CNN GLENTS is its focus on evaluating not only linguistic competence but also the examinees’ awareness of global topics, making it suitable for those interested in international media, journalism, or globally relevant industries (Asahi Press, 2024). Nonetheless, detailed research on CNN GLENTS itself remains limited, especially in comparison to established tests like the TOEIC and GTEC. In addition, the CNN GLENTS was discontinued in March 2024, making it no longer available. However, comparing a test that places greater emphasis on background knowledge with other assessments is still valuable for identifying effective uses of tests in English teaching contexts.

2.2 Evaluations of Tests

To assess the effects of educational programs comprehensively, it is recommended to combine subjective methods, such as surveys, with objective evaluations, such as tests (Bachman & Palmer, 2010). In this context, researchers employ various tests to measure English proficiency, and understanding the relationships between them can enhance the accuracy of program evaluation. For instance, Kanzaki (2021) analyzed the TOEIC and International English Language Testing System test results of 84 university students, revealing significant correlations between the scores of corresponding skills across both tests, with even stronger correlations observed in the total scores. Building on this, Sugiyama et al. (2022) examined the relationships among the GTEC-Academic, TOEIC, and Global Proficiency Skills-Academic. They found that the experimental group

receiving 1 year of transmission-oriented English instruction scored significantly higher in English writing compared to those who did not. However, many aspects remain unclear, such as the long-term retention of skills gained through different instructional methods and how these correlations may vary across different demographic groups.

2.3 The Role of Background Knowledge

Many studies have demonstrated that possessing background knowledge plays a crucial role in understanding and acquiring a foreign language (Alderson, 2000; Hauptman, 2000). However, some studies suggest that background knowledge can have a negative impact on the acquisition of new information (Alexander & Judy, 1988; Dochy et al., 1999). Carrel (1983) compared the role of background knowledge in reading between native English speakers and non-native speakers, revealing that native speakers were more adept at utilizing their background knowledge, whereas non-native speakers struggled to do so. Moreover, the relationship between background knowledge and skills such as listening and speaking remains largely unexplored. Clarifying these relationships could lead to better methods of presenting information and, ultimately, to more effective language instruction. In PEP, students gather previous research before conducting experiments aimed at new discoveries. During this process, it is essential for them to collect sources impartially, including those that differ from their own perspectives, and to understand and present these findings to others without bias. Therefore, comparing CNN GLENTS, which incorporates background knowledge into its assessments, with other tests holds significant value.

3. Method

3.1 Participants

The participants in this study were 31 students enrolled at a large private university in Japan, consisting of 14 first-year students and 17 second-year students. The first-year and second-year students were in separate classes, both taught by one of the authors. We applied no specific selection criteria to narrow down the participants. They attend two English classes per week: one focused on PEP, and the other a Skill Workshop designed to improve fundamental English skills, including reading, listening, speaking, writing, and grammar. The PEP course spans from the first-year spring semester to the second-year fall semester, with instructors referring to each semester's classes as Project English 1–4 (hereafter, P1–P4). For students in the College of Life Sciences and Pharmaceutical Sciences, Junior Project 1 (hereafter, JP1) in the first semester of their third year is also a required course. In P1 and P2, students become familiar with academic presentations, whereas in P3, they engage in group activities such as debates and panel discussions. The goal of P4 is to enable students to write an academic paper of approximately 15 pages, along with giving oral presentations. In JP1, students focus on specialized topics, receive feedback from professors and other faculty members in their respective fields, and deliver a poster presentation in the style of an international academic conference.

3.2 Procedures

We conducted the experiment in the AY2023. Table 1 shows the content of each test and the approximate timing of the exams. The participants took each of the three tests twice. Because they took the tests online individually, there was a slight variation in the exact dates, but we scheduled all tests to be taken approximately 8 months apart. Since the purpose of the study was to assess the program's outcomes, no test-specific practice was provided. The students took the tests while participating in the regular program activities.

Table 1. Summary of the Tests

Test	Skills Assessed	1st Session	2nd Session
CNN GLENTS	Listening, Reading, Global Knowledge	April 14, 2023	December 8, 2023
GTEC-Academic	Listening, Reading, Speaking, Writing	Early April–May 2023	Late December 2023–January 2024
TOEIC IP	Listening, Reading	March 9–23, 2023	December 6–8, 2023

To examine the relationship between each test and its sections, we conducted the following analyses. First, we compared mean scores for each section of the tests to assess statistical significance; for within-examinee comparisons we used Excel's paired t-test, whereas for between-examinee comparisons we used Excel's independent t-test assuming equal variances. Second, we calculated correlation coefficients for any two items across all test sections. Third, we assessed score improvements by comparing scores from the first and second attempts for each test section, and calculated correlation coefficients for any two items. Fourth, given the significant differences in test scores between the two classes (Class A, consisting of first-year students, and Class

B, consisting of second-year students), we isolated data from Class B to compute the correlation coefficients as outlined in the second step. Finally, in the second TOEIC test, we categorized those who scored 600 or above as 'high achievers,' and we conducted the analysis on these students, primarily focusing on their CNN GLENTS results.

4. Results

4.1 Results of Each Test

We present the mean scores for each section of the tests, along with a report of the items that exhibited statistically significant differences ($p < 0.05$, two-tailed). Class A consists of 14 first-year students, while Class B consists of 17 second-year students.

4.1.1 TOEIC Results

Table 2 shows the results of TOEIC tests. Both Class A and Class B scored higher on their second attempt in LR and total scores. In the comparison between Class A and Class B, we observed statistically significant differences in the first listening, second listening, first reading, first total score, and second total score ($p < 0.05$, two-tailed).

Table 2. TOEIC Results

	Listening (1st)	Listening (2nd)	Reading (1st)	Reading (2nd)	Total (1st)	Total (2nd)
Total	298.5	308.3	252.5	259.0	551.0	567.3
Class A	265.0	274.1	215.5	228.6	480.5	502.7
Class B	326.9	337.3	283.8	284.6	610.8	621.9

4.1.2 GTEC Results

Table 3 shows the results of GTEC tests. In both classes, the scores for listening and reading decreased slightly on the second attempt, whereas the scores for speaking and writing showed a slight increase. A comparison of the first and second tests revealed statistically significant differences in overall speaking scores. Additionally, in the comparison between Class A and Class B, we observed statistically significant differences in the first speaking, second speaking, first writing, first total score, and second total score.

Table 3. GTEC Results

	Listening (1st)	Listening (2nd)	Reading (1st)	Reading (2nd)	Speaking (1st)	Speaking (2nd)	Writing (1st)	Writing (2nd)	Total (1st)	Total (2nd)
Total	126.6	124.8	118.5	116.8	119.8	125.8	123.3	126.8	488.2	494.2
Class A	118.4	116.1	116.3	116.7	112.1	116.3	110.4	117.5	457.1	466.5
Class B	133.6	132.2	120.3	116.8	126.2	133.8	134.3	134.7	514.5	517.6

4.1.3 CNN GLENTS Results

Table 4 shows the scores of the first and second rounds of the CNN GLENTS test. In both classes, the scores for international studies were higher in the second round than in the first, whereas the listening scores were higher in the first round for both classes. We observed a statistically significant difference in the reading scores of Class B between the first and second rounds, as well as between the reading scores of Classes A and B in the second test.

Table 4. CNN GLENTS Results

	Listening (1st)	Listening (2nd)	Reading (1st)	Reading (2nd)	International Studies (1st)	International Studies (2nd)	Total (1st)	Total (2nd)
Total	26.0	24.4	17.4	20.7	6.9	8.2	50.3	53.3
Class A	25.1	23.8	18.0	17.5	7.1	7.6	50.2	48.9
Class B	26.8	24.9	16.9	23.4	6.8	8.6	50.5	56.9

4.2 Correlations Across Tests

Table 5 shows the correlations between each item of CNN GLENTS and the items of GTEC and TOEIC. Additionally, Table 6 displays the correlations between the items of TOEIC and GTEC. Together, these two tables present all the correlation coefficients between any two selected items. When examining the correlations

between CNN GLENTS and other tests, most showed either no correlation or weak correlations. Additionally, there were 14 negative correlations between CNN GLENTS and the items of GTEC, and six negative correlations with the items of TOEIC. Notably, in the second round of international studies, 11 out of 16 items (68.8%) exhibited negative correlations. The correlations between GTEC and TOEIC were generally higher compared to those involving CNN GLENTS, with a particularly high correlation coefficient of .72 recorded between the listening components of the second round of both tests. Comparing the first and second tests, the correlations between international studies and each component were lower in the second test, except for TOEIC reading.

Table 5. Correlations Between CNN GLENTS and GTEC and TOEIC

		Listening (1st)	Reading (1st)	International Studies (1st)	Total (1st)	Listening (2nd)	Reading (2nd)	International Studies(2nd)	Total (2nd)
CNN GLENTS Score	Listening (1st)	1	.54	.18	.81	.08	.05	.43	.19
	Reading (1st)	.54	1	.47	.89	.30	.09	.11	.23
	International Studies (1st)	.18	.47	1	.59	-.17	-.14	.02	-.16
	Total (1st)	.81	.89	.59	1	.14	.03	.27	.16
	Listening (2nd)	.08	.30	-.17	.14	1	.52	-.11	.77
	Reading (2nd)	.05	.09	-.14	.03	.52	1	.15	.90
	International Studies (2nd)	.43	.11	.02	.27	-.11	.15	1	.32
	Total (2nd)	.19	.23	-.16	.16	.77	.90	.32	1
GTEC Score	Listening (1st)	.43	.34	.34	.47	.25	.11	-.20	.13
	Reading (1st)	.20	.37	.21	.33	.32	.20	-.31	.19
	Speaking (1st)	.29	.16	.16	.27	.30	.26	-.10	.27
	Writing (1st)	.41	-.11	-.08	.13	-.01	.14	.10	.11
	Total (1st)	.41	.27	.22	.40	.28	.21	-.18	.21
	Listening (2nd)	.44	.32	.10	.41	.54	.33	.00	.46
	Reading (2nd)	.18	.26	.00	.22	.39	.33	-.31	.30
	Speaking (2nd)	.44	.16	.07	.32	.18	.35	.13	.33
TOEIC Score	Writing (2nd)	.44	-.21	.06	.13	-.10	-.05	-.12	-.11
	Total (2nd)	.53	.19	.08	.38	.38	.34	-.13	.35
	Listening (1st)	.22	.32	.24	.33	.41	.32	-.08	.37
	Reading (1st)	.20	.04	-.02	.11	.29	.62	.06	.53
	Total (1st)	.23	.19	.11	.24	.38	.52	-.01	.50
	Listening (2nd)	.21	.29	-.02	.24	.62	.58	-.16	.61
	Reading (2nd)	.27	.31	.12	.32	.54	.57	.14	.64
	Total (2nd)	.27	.33	.06	.31	.65	.64	-.01	.69

Table 6. Correlations Between TOEIC and GTEC

		Listening (1st)	Reading (1st)	Total (1st)	Listening (2nd)	Reading (2nd)	Total (2nd)
GTEC Score	Listening (1st)	.51	.24	.40	.66	.35	.55
	Reading (1st)	.39	.31	.39	.54	.43	.53
	Speaking (1st)	.60	.45	.58	.62	.40	.56
	Writing (1st)	.31	.40	.40	.35	.38	.40
	Total (1st)	.56	.41	.53	.68	.48	.64
	Listening (2nd)	.59	.36	.52	.72	.44	.64
	Reading (2nd)	.26	.32	.33	.53	.39	.51
	Speaking (2nd)	.51	.48	.55	.49	.32	.45
	Writing (2nd)	.21	.24	.25	.30	.13	.23
Total (2nd)	.56	.49	.58	.74	.46	.66	

4.3 Comparison of Score Improvements and Correlation Analysis Between Sections

We compared the scores from the first and second rounds across the sections of each test to calculate the score improvement, and determined the correlation coefficients between any two selected items. The results of this analysis did not reveal any noteworthy correlations.

4.4 Analysis of Correlation Coefficients in Class B

In the comparison of average scores by class, there were many tests where the difference between Class A and Class B was large, with Class B generally having higher average scores. Therefore, we extracted data from Class B, and calculated the correlation coefficients between any two items across all sections of all tests.

Table 7 presents the correlation coefficients between the individual items of CNN GLENTS and those of GTEC and TOEIC. Table 8, meanwhile, illustrates the correlations between the items of TOEIC and GTEC. Collectively, these tables provide a comprehensive view of the correlation coefficients across all item pairs.

In the second GTEC test, the writing section showed a negative correlation with all other sections. Additionally, the speaking section in the second GTEC test also had low correlation coefficients with other sections. We observed a strong correlation of .82 between the listening section of the second GTEC test and the listening section of the second TOEIC test. The listening section of the second TOEIC test also showed a strong correlation with the writing section of the second GTEC test. Further, the second listening sections of the TOEIC, GTEC, and CNN GLENTS tests all demonstrated strong correlations.

Table 7. Correlations Between CNN GLENTS and GTEC and TOEIC in Class B

		Listening (1st)	Reading (1st)	International Studies (1st)	Total (1st)	Listening (2nd)	Reading (2nd)	International Studies (2nd)	Total (2nd)
CNN GLENTS Score	Listening (1st)	1	.53	.24	.79	-.02	-.26	.22	-.12
	Reading (1st)	.53	1	.52	.91	.50	.15	.20	.41
	International Studies (1st)	.24	.52	1	.65	-.18	-.07	.11	-.11
	Total (1st)	.79	.91	.65	1	.21	-.05	.23	.14
	Listening (2nd)	-.02	.50	-.18	.21	1	.51	-.33	.77
	Reading (2nd)	-.26	.15	-.07	-.05	.51	1	.02	.90
	International Studies (2nd)	.22	.20	.11	.23	-.33	.02	1	.08
	Total (2nd)	-.12	.41	-.11	.14	.77	.90	.08	1
GTEC Score	Listening (1st)	.31	.38	.52	.47	.21	-.16	-.62	-.14
	Reading (1st)	-.06	.29	.38	.23	.44	.17	-.70	.17
	Speaking (1st)	.18	.39	.34	.38	.21	-.02	-.58	-.04
	Writing (1st)	.47	-.07	.17	.21	-.27	-.39	-.39	-.49
	Total (1st)	.23	.33	.46	.40	.24	-.08	-.70	-.10
	Listening (2nd)	.26	.40	.05	.34	.66	.18	-.59	.32
	Reading (2nd)	.07	.04	-.26	-.01	.60	.48	-.59	.47
	Speaking (2nd)	.22	.29	.34	.34	-.07	-.04	-.23	-.12
TOEIC Score	Writing (2nd)	.29	-.33	.16	-.01	-.47	-.54	-.51	-.72
	Total (2nd)	.30	.11	.05	.20	.33	.07	-.76	.03
	Listening (1st)	.19	.63	.32	.50	.68	.32	-.32	.48
	Reading (1st)	.12	.06	.08	.11	.25	.66	-.10	.52
	Total (1st)	.18	.36	.22	.33	.51	.62	-.23	.60
	Listening (2nd)	.03	.42	.04	.25	.76	.41	-.51	.54
	Reading (2nd)	.33	.55	.37	.53	.54	.51	.00	.61
	Total (2nd)	.21	.55	.24	.45	.72	.52	-.27	.65

Table 8. Correlations Between TOEIC and GTEC in Class B

		Listening (1st)	Reading (1st)	Total (1st)	Listening (2nd)	Reading (2nd)	Total (2nd)
GTEC Score	Listening (1st)	.47	.03	.26	.62	.30	.51
	Reading (1st)	.68	.37	.60	.68	.42	.61
	Speaking (1st)	.43	.16	.32	.53	.21	.40
	Writing (1st)	.23	.18	.24	.02	-.05	-.02
	Total (1st)	.57	.21	.44	.62	.31	.51
	Listening (2nd)	.62	.15	.42	.82	.50	.73
	Reading (2nd)	.43	.41	.50	.74	.47	.67
	Speaking (2nd)	.30	.15	.26	.10	-.01	.05
	Writing (2nd)	-.12	-.08	-.11	-.06	-.17	-.13
	Total (2nd)	.46	.26	.41	.67	.34	.55

4.5 Analysis of Participants with High TOEIC Scores

We classified participants who scored 600 points or higher in the second TOEIC test as ‘high-level.’ We conducted the analysis focusing particularly on the CNN GLENTS results for these students. The subjects consisted of a total of 10 students: two first-year students and eight second-year students. Table 9 shows the correlations with the second round of each test, and Table 10 presents the correlations with the improvement from the first to the second round of the TOEIC and GTEC tests.

When examining the correlations between the second CNN GLENTS test and the score growth from the first to the second round of each test, we observed strong correlations among the listening, reading, and international studies sections of CNN GLENTS. In contrast, there were either no correlations or negative correlations with other tests. Notably, all correlations between each section of TOEIC and the reading and international studies sections were negative.

Table 9. Correlations With the Second Round of Each Test

		Listening	Reading	International Studies	Total
CNN GLENTS Score	Listening	1	.47	-.52	.72
	Reading	.47	1	.08	.93
	International Studies	-.52	.08	1	.03
	Total	.72	.93	.03	1
GTEC Score	Listening	.38	.01	-.57	.06
	Reading	.67	.31	-.54	.42
	Speaking	-.08	-.11	-.28	-.19
	Writing	-.53	-.83	-.26	-.89
TOEIC Score	Total	.19	-.27	-.69	-.25
	Listening	.56	.30	-.71	.31
	Reading	.22	.63	.03	.54
	Total	.59	.65	-.54	.60

Table 10. Correlations With the Growth from the First to the Second Round of TOEIC and GTEC

		Listening	Reading	International Studies	Total
CNN GLENTS Score Growth	Listening	.87	.54	-.54	.68
	Reading	.37	.87	.17	.81
	International Studies	-.08	-.02	.71	.13
	Total	.69	.74	-.06	.84
GTEC Score Growth	Listening	.41	.40	.37	.58
	Reading	.19	.21	.41	.35
	Speaking	-.05	.15	.43	.18
	Writing	-.43	-.52	-.25	-.64
TOEIC Score Growth	Total	.16	.18	.47	.33
	Listening	.30	-.03	-.56	-.00
	Reading	.24	-.20	-.26	-.07
	Total	.31	-.14	-.46	-.04

5. Discussion and Future Study

The findings from this study provide insight into the relationships among the CNN GLENTS, GTEC, and TOEIC tests, focusing on the correlations between individual items across multiple testing rounds. The analysis reveals that the listening and reading sections of GTEC in the second round showed higher correlation coefficients with the corresponding sections of TOEIC, possibly due to increased familiarity with the test formats. As participants became more accustomed to the structure and style of the tests, their anxiety levels likely decreased, leading to more consistent performance across both assessments (Horwitz, 2001). Conversely, the lower correlations observed in speaking and writing may reflect less familiarity with these productive skills, highlighting the complexity of these tasks.

As Raimes (1983) emphasized, writing is a multifaceted skill that involves numerous cognitive components, such as organization, grammar, and vocabulary, which may require different teaching and assessment strategies. This complexity could explain the weaker correlations between GTEC's speaking and writing sections and the TOEIC in the second round, given these skills likely require more practice and tailored instruction for consistent improvement. The findings suggest that improvements in productive skills may not align as directly with standardized test formats as receptive skills like listening and reading.

CNN GLENTS generally exhibited weak or negative correlations with the other tests, indicating that it measures constructs distinct from those assessed by GTEC and TOEIC. The second round of the international studies section in CNN GLENTS showed particularly high rates of negative correlations (68.8%), suggesting a focus on skills that differ significantly from the language-specific proficiencies measured by the other tests. This finding aligns with earlier studies on test construct divergence (Bachman & Palmer, 2010). It is possible that CNN GLENTS places a greater emphasis on background knowledge, which may cause participants to rely less on language performance and more on their existing knowledge base when answering questions. It is also possible that the use of novel topics in the tests, distinct from what participants had encountered before, made background knowledge a disadvantage. Further, negative correlations were generally larger for listening than for reading. This may be because, although participants could reread texts in the reading sections, they could not do so in listening sections where they had to respond immediately to what they heard (Hirai, 1999). These observations suggest that when delivering oral presentations, presenters should consider that the audience's background knowledge might negatively impact their understanding. It is important to provide more detailed explanations that take this factor into account.

These results suggest that different language assessments may prioritize varying skill sets, which has implications for selecting tests in educational settings. Understanding these differences can help educators better align test choices with the specific needs of their students. For example, teachers might use CNN GLENTS for students who need to engage with content-based assessments, whereas TOEIC and GTEC could be more suitable for evaluating general language proficiency.

One limitation of this study is the small sample size, which may limit the generalizability of the findings. Additionally, we did not collect qualitative data, such as interviews, that could provide deeper insight into the reasons behind participants' performance. Future researchers should investigate how background knowledge impacts listening and speaking performance, and how test familiarity influences score consistency over longer periods.

6. Conclusion

In this study we examined the relationships between the items of multiple tests used to measure English proficiency. The findings revealed that even tests designed to assess similar abilities can show both high and low correlations among their items. Therefore, it is important not to rely solely on a single test to fully gauge a student's English proficiency, but rather to consider multiple approaches. Additionally, although further validation is needed, it is noteworthy that possessing greater background knowledge may negatively impact performance on tests measuring English proficiency. It is also essential to explore how background knowledge can be used effectively in language assessments and learning.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732935>
- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58, 375-404. <https://doi.org/10.3102/00346543058004375>
- Asahi Press. (2024). *CNN Global English Testing System overview*. <https://www.asahipress.com/special/glents/>
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press. <https://doi.org/10.1016/B978-0-08-044894-7.00263-3>
- Benesse Corporation. (n.d.). *GTEC test overview*. <https://www.benesse.o.jp/gtec/>
- Butler, Y.G., & Iino, M. (2005). Current Japanese Reforms in English Language Education: The 2003 "Action Plan". *Lang Policy*, 4, 25-45. <https://doi.org/10.1007/s10993-004-6563-5>
- Carrell, P. L. (1983). Three components of background knowledge in reading comprehension 1. *Language Learning*, 33(2), 183-203. <https://doi.org/10.1111/j.1467-1770.1983.tb00534.x>
- Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, 69, 145-186. <https://doi.org/10.3102/00346543069002145>
- Educational Testing Service. (2024). *TOEIC test: Test content and scoring*. <https://www.ets.org/toeic>
- Hauptman, P. C. (2000). Some hypotheses on the nature of difficulty and ease in second language reading: An application of schema theory. *English Language Annual*, 33(6), 622-631. <https://doi.org/10.1111/j.1944-9720.2000.tb00931.x>
- Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *Modern Language Journal*, 83(3), 367-384. <https://doi.org/10.1111/0026-7902.00028>
- Horwitz, E. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112-126. <https://doi.org/10.1017/S0267190501000071>
- Kanzaki, M. (2021). Correlations between IELTS and TOEIC scores. In P. Clements, R. Derrah, & P. Ferguson (Eds.), *Communities of teachers & learners* (pp. 397-406). JALT. <https://doi.org/10.37546/JALTPCP2020-49>
- Kim, M., Smith, W. Z., & Chin, T. Y. (2017, February). *Validation and linking scores for the Global Test of English Communication: White paper*. <https://www.benesse.co.jp/gtec/schoolofficials/research/pdf/doc-2016-02.pdf>
- Ministry of Education, Culture, Sports, Science and Technology of Japan. (2002, July 12). "Eigo ga tsukaeru nihonjin" no ikusei no tameno senryakukousou no sakutei nit suite (On the formulation of a strategic concept for the development of English-speaking Japanese) [in Japanese]. https://www.mext.go.jp/b_menu/shingi/chousa/shotou/020/sesaku/020702.htm
- Powers, D. E., & Powers, A. (2015). The incremental contribution of TOEIC® Listening, Reading, Speaking, and Writing tests to predicting performance on real-life English language tasks. *Language Testing*, 32(2), 151-167. <https://doi.org/10.1177/0265532214551855>

- Raimes, A. (1983). *Techniques in teaching writing*. Oxford University Press.
- Sugiyama, K., Yamanaka, T., & Odagiri, K. (2022). What are the effects of project-based English curriculum on the development of learners' competencies? A case study of a Japanese university English language program. *English Language Teaching*, 15(12), 60-70. <https://doi.org/10.5539/elt.v15n12p60>
- Yamanaka, T., Kimura, S., Yamashita, M., & Kondo, Y. (2021). *Purojekuto hasshingata eigo puroguramu: jibun jikuwo kitaeru "oshienai" kyouiku (Project-based English Program: Train your own axis "without teaching")* [in Japanese]. Kitaohji Shobo.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).