

Longitudinal Analyses of English Proficiency Testing Results Collected From First-Year Undergraduate Students

Daisuke YAMAZAKI

Centre for Liberal Arts and Sciences, Faculty of Engineering

Abstract

This paper sets out to investigate whether there is any improvement in English proficiency demonstrated by the first-year undergraduate students at Toyama Prefectural University in Japan, by analysing the results of a standardized English examination (*TOEIC® Bridge Listening and Reading Tests*) conducted twice, once in April 2021 as a pretest and once in December 2021 as a posttest. In order to achieve the main purpose of this study, two research questions are formulated. (1) Is there a significant difference between the mean scores of the pretest and posttest? (2) Is there some kind of development in English skills and abilities that the test subjects have attained between the pretest and posttest? As a result, regarding the first question, it may fairly be presumed that there was no statistically significant difference between the mean scores of the pretest and posttest. With regard to the second question, it can be inferred, in five out of eight items related to the percentage of correct answers per skill called *Abilities Measured*, that there might have been some improvements between the pretest and posttest, especially in the aspects of vocabulary and listening for the main idea or stated fact. It is hoped that some standardized proficiency tests to measure and track English skills and abilities that learners have acquired are appropriately implemented at school on a regular basis, and the results of the examinations are scrutinized closely and properly, for the sake of employing the eventual outcome effectively in educational settings.

Keywords: English proficiency test, first-year undergraduate students, longitudinal analyses, pretest and posttest, percentage of correct answers per skill

1. Background

To visualize learning outcome performed by students with solid data could be indispensable under a policy to assure the internal quality at a tertiary institution. As a part of the institutional research (IR) on university, some standardized tests to measure English proficiency that students achieve may have seemingly been conducted throughout the nation.

At Toyama Prefectural University (hereinafter TPU), the author as the director for English education reform has been devotedly committed to implement some English proficiency tests for the undergraduate students since 2015. During the initial stage, for the sake of expediency and other practical reasons, *A.C.E. Placement (Assessment of Communicative English-Placement)* called *English Placement Test* developed by the Association for English Language Proficiency Assessment (ELPA) was utilized to

measure the levels of listening and reading proficiency in the English language. Yamazaki (2019) analysed the results of the English tests collected from the separate first-year engineering undergraduate students at TPU, comparing between the data for April 2017 and that for April 2018.

In recent years, a growing globalization in both foreign-affiliated and domestic enterprises may have been wielding some influence over the motivation of English learners, especially job seekers and those who stand at the crossroads of their career, to boost the scores of some English qualification tests, and there is no exception in that a large number of the TPU students are eager to improve their scores of TOEIC, or, in other words, one of the most well-known English examinations among university students in Japan. Just for reference, as a matter of fact, any applicants for admission to the engineering graduate programme at TPU are required to submit a TOEIC

Official Score Certificate or a *TOEFL Examinee Score Report/Test Taker Score Report* at the time of their application.

Moreover, the rampant Coronavirus (COVID-19) situations worldwide might have been forcing the administrators and operational managers who are in charge of implementing some standardized English examinations at school to make careful consideration for countermeasures in case of emergency or any unforeseen circumstances. In actuality, from the viewpoint of tracking year-to-year changes for a pure comparison, it is obviously reasonable to set the date of an exam implementation at the same time each year, so it is necessary to have an alternative measurement to conduct such standardized tests at a designated period without any delay or postponement. As for some test schemes of TOEIC (Institutional Programme), since April 2020, it has become possible for its test takers to sit for the exam on the Internet remotely, which means that the online version can be carried out even if a curfew or other binding orders not to go out of home is imposed on the city or area affected by the raging virus. In addition, TPU has been a paid institutional member of the Institute for International Business Communication (hereinafter IIBC), the governing body for implementing TOEIC in Japan, and one of the membership privileges enables TPU students to obtain the benefit of taking some TOEIC tests with a discounted price. After considering the main factors mentioned above, at TPU, TOEIC has become a leading candidate which can be introduced to measure English proficiency that the students demonstrate.

In both April 2021 as a pretest and December 2021 as a posttest, *TOEIC® Bridge Listening and Reading Tests* (hereinafter TOEIC Bridge L&R) were administered, targeting all the first-year undergraduate students who belong to the nursing and engineering departments at TPU.

The main purpose of this study is to investigate whether there is any improvement in English proficiency that the TPU students as the test subjects of this study have acquired between the pretest and posttest. In order to accomplish the objective, two research questions are formulated. (1) Is there a significant difference between the mean scores of the pretest and posttest? (2) Is there some kind of development in English skills and abilities that the test subjects have attained between the pretest and posttest?

2. Method

Subjects

All of the first-year undergraduate students at TPU in Academic Year 2021, i.e., four hundred eighty students (121 nursing and 359 engineering members), were supposed to take TOEIC Bridge L&R twice. Among them, in actuality, four hundred seventy-five students took the examination in April as a pretest, whereas four hundred nineteen subjects sat for the test in December as a posttest. However, in order to make a pure comparison between the results of the pretest and posttest possible, four hundred seventeen students who took both the pretest and posttest were after all targeted as the test subjects of this study.

Materials

All the test subjects took TOEIC Bridge L&R twice during a specified interval of time, while the tests differed in content between the pretest and posttest. Concerning the level of the English tests, they are described as “for beginning to lower-intermediate learners”, according to IIBC (n.d.) a. The details of the test contents are shown in Table 1, which are basically extracted from the English-version website of IIBC (n.d.) b.

Table 1. *Test Content of TOEIC Bridge L&R*

Part	Content	No. of Questions
Listening (approximately 25 minutes, 50 questions)		
Part 1	Four Pictures	6
Part 2	Question-Response	20
Part 3	Conversations	10
Part 4	Talks	14
Reading (35 minutes, 50 questions)		
Part 1	Sentence Completion	15
Part 2	Text Completion	15
Part 3	Reading Comprehension	20

Procedure

For the engineering students, the pretest was conducted during the period of the orientation session for newly incoming members on the 5th of April 2021, i.e., the day before the TPU entrance ceremony. And for the nursing students, it was carried out during the first session of the English II course on the 9th of April 2021.

With regards to the posttest, it was on the 10th of December 2021 for both the nursing and engineering

students. The nursing subjects took the English examination during the ninth session of the English IV course, while the engineering students took it outside class hours in the afternoon of the same day.

3. Results

The entire results of TOEIC Bridge L&R collected from the test subjects taking both the pretest in April 2021

and the posttest in December 2021 are shown in Table 2. The mean total score of the pretest was 67.68 out of 100, whereas 67.20 was that of the posttest. In Table 2, the sectional results of both Listening and Reading are also included in each. The listening average score added to the reading counterpart becomes the total score. For instance, in the case of the pretest, 30.01 out of 50 as the listening mean score plus 37.67 out of 50 as the reading one equals 67.68 as the total average score.

Table 2. Overall Results of TOEIC Bridge L&R Collected From Test Subjects Taking Both Pretest and Posttest

Test	Score Classification	<i>n</i>	Mean	<i>SD</i>	<i>SE</i>	Minimum	Maximum
Pretest	Total	417	67.68	10.45	0.51	36	94
Posttest		417	67.20	10.64	0.52	33	99
Pretest	Listening	417	30.01	5.65	0.28	15	46
Posttest		417	29.60	6.10	0.30	15	50
Pretest	Reading	417	37.67	6.17	0.30	16	50
Posttest		417	37.60	5.79	0.28	15	50

The data concerning the percentage of correct answers per skill called *Abilities Measured*, which was

reckoned and provided in the *Score Roster* by IIBC, are shown in Table 3.

Table 3. Percentage of Correct Answers per Skill Obtained by Test Subjects Taking Both Pretest and Posttest

Test	Abilities Measured	<i>n</i>	Mean	<i>SD</i>	<i>SE</i>	Minimum	Maximum
Pretest	L1	417	70.64%	14.16	0.69	15%	100%
Posttest		417	65.79%	13.73	0.67	20%	95%
Pretest	L2	417	68.82%	13.32	0.65	20%	97%
Posttest		417	67.79%	13.32	0.65	23%	97%
Pretest	L3	417	57.39%	17.18	0.84	0%	100%
Posttest		417	70.34%	16.92	0.83	7%	100%
Pretest	L4	417	59.44%	16.82	0.82	0%	100%
Posttest		417	73.07%	16.77	0.82	6%	100%
Pretest	R1	417	63.62%	14.64	0.72	21%	100%
Posttest		417	79.32%	14.91	0.73	0%	100%
Pretest	R2	417	67.70%	16.97	0.83	15%	100%
Posttest		417	68.52%	17.56	0.86	0%	100%
Pretest	R3	417	65.87%	18.23	0.89	7%	100%
Posttest		417	75.65%	17.53	0.86	0%	100%
Pretest	R4	417	66.39%	17.31	0.85	10%	100%
Posttest		417	73.18%	16.47	0.81	0%	100%

Concrete details about each item of *Abilities Measured* in listening and reading are explained in the

English-version website of IIBC (n.d.) c. Basically, the following descriptions are extracted from the webpage.

Abilities Measured (percentage of correct answers per skill)

Listening

L1 → *Appropriate Response*: how well you identified an appropriate spoken response to a speaker in a brief spoken exchange.

L2 → *Short Dialogue or Conversation*: how well you understood a short dialogue or conversation spoken slowly.

L3 → *Short Monologue*: how well you understood a short talk spoken slowly by one speaker.

L4 → *Listening for Main Idea or Stated Fact*: how well you understood the main idea or stated facts in short conversations and talks.

Reading

R1 → *Vocabulary*: how well you understood words or short phrases in simple written sentences.

R2 → *Grammar*: how well you understood the form, meaning, and use of simple grammatical structures in simple written sentences.

R3 → *Reading for Main Idea or Stated Fact*: how well you understood the main idea or stated facts in short written texts.

R4 → *Short Informational Written Texts*: how well you understood short, informational, descriptive, and expository written texts.

(Note that the labels *L1*, *L2*, *L3*, *L4*, *R1*, *R2*, *R3*, *R4* are appended above by the author after confirming with IIBC.)

By looking at the percentage of correct answers per skill, it may be possible to grasp the English abilities that the test subjects have attained. For example, as shown in Table 3, the mean value of the percentage for *R1* concerning vocabulary was 63.62% in the pretest, whereas the value increased to 79.32% in the posttest. Supposedly, there is some possibility that the skill in vocabulary of the test subjects might have been more or less improved in a sense.

4. Discussion

(1) Comparing the Mean Scores of TOEIC Bridge L&R

As shown in Table 2, it seems that the mean value of the total score slightly decreased in the posttest compared to that in the pretest. Based on all the sets of data, some statistical analyses were appropriately conducted to substantiate the difference between the pretest and posttest. Tests of statistical significance were conducted using computer software IBM® SPSS® Statistics Version 26. As a result, there was no statistical significance at the $p < .05$ level of confidence, concerning the difference in the mean values of the TOEIC Bridge L&R scores between the pretest and posttest (see Table 4). Analyses of the total score resulted in the following statistics: $t(832) = 1.159$, $p = .247$, $d = 0.05$, 95% CI [-0.33, 1.29]. Therefore, it might be inferred that there was no statistically significant difference between the mean scores of the pretest and posttest.

Table 4. *Statistical Analyses of TOEIC Bridge L&R Results Collected From Test Subjects Taking Both Pretest and Posttest*

Test	Score Classification	<i>n</i>	Mean	<i>SD</i>	Mean Difference	$t(832)$	<i>p</i>	Effect Size <i>d</i>	95% CI	
									<i>LL</i>	<i>UL</i>
Pretest	Total	417	67.68	10.45	0.48	1.159	.247	0.05	-0.33	1.29
Posttest		417	67.20	10.64						
Pretest	Listening	417	30.01	5.65	0.41	1.500	.134	0.07	-0.13	0.94
Posttest		417	29.60	6.10						
Pretest	Reading	417	37.67	6.17	0.07	.276	.782	0.01	-0.44	0.58
Posttest		417	37.60	5.79						

(2) Development in English Skills

According to a close observation of Table 3, it was found that the percentage of two items in *Abilities Measured* deteriorated in the posttest compared to the pretest, while the other six percentages more or less made

a progress between the pretest and posttest. Then, some statistical analyses were appropriately conducted to substantiate the difference between the two tests. As a result, in terms of the six items in *Abilities Measured*, i.e., *L1* (Appropriate Response), *L3* (Short Monologue), *L4*

(Listening for Main Idea or Stated Fact), *R1* (Vocabulary), *R3* (Reading for Main Idea or Stated Fact), and *R4* (Short Informational Written Texts), statistically significant differences were found at the $p < .001$ level of confidence, in comparison of the mean values related to the correct answer percentage per skill between the pretest and posttest (see Table 5). Among the six items whose differences are statistically significant between the prior and posterior tests, *L1* was the only one whose percentage had decreased in the posttest compared to the pretest, while, concerning the other five aspects, the percentage in the posttest had increased compared to the one in the pretest.

In particular, *R1* (Vocabulary) showed the largest difference in mean values for reading between the pretest

and posttest (see Table 5, also Figure 1 for the median and distribution), whereas *L4* (Listening for Main Idea or Stated Fact) was uppermost for listening (see Table 5, also Figure 2 for the median and distribution). From these results, it can be inferred that the test subjects might have somewhat reinforced their vocabulary and a part of their listening skills after entering TPU.

On the other hand, as can be seen in Table 5, *R2* (Grammar) and *L2* (Short Dialogue or Conversation) did not show much difference between the pretest and posttest. Therefore, the results can imply that enhancing the skills of these aspects, i.e., grammar and listening to a short dialogue or conversation, might help trigger an improvement in TOEIC scores of the test subjects.

Table 5. Statistical Analyses of Correct Answer Percentage per Skill in TOEIC Bridge L&R Taken by Test Subjects

Test	Abilities Measured	<i>n</i>	Mean	<i>SD</i>	Mean Difference	<i>t</i> (832)	<i>p</i>	Effect Size <i>d</i>	95% CI	
									<i>LL</i>	<i>UL</i>
Pretest	L1	417	70.64%	14.16	4.85	7.110	.000	0.35	3.50	6.18
Posttest		417	65.79%	13.73						
Pretest	L2	417	68.82%	13.32	1.03	1.605	.109	0.08	-0.23	2.29
Posttest		417	67.79%	13.32						
Pretest	L3	417	57.39%	17.18	12.95	-12.858	.000	0.76	-14.92	-10.97
Posttest		417	70.34%	16.92						
Pretest	L4	417	59.44%	16.82	13.63	-14.461	.000	0.81	-15.48	-11.78
Posttest		417	73.07%	16.77						
Pretest	R1	417	63.62%	14.64	15.70	-19.362	.000	1.06	-17.30	-14.11
Posttest		417	79.32%	14.91						
Pretest	R2	417	67.70%	16.97	0.82	-0.902	.368	0.05	-2.59	0.96
Posttest		417	68.52%	17.56						
Pretest	R3	417	65.87%	18.23	9.78	-10.264	.000	0.55	-11.65	-7.91
Posttest		417	75.65%	17.53						
Pretest	R4	417	66.39%	17.31	6.79	-8.074	.000	0.40	-8.44	-5.14
Posttest		417	73.18%	16.47						

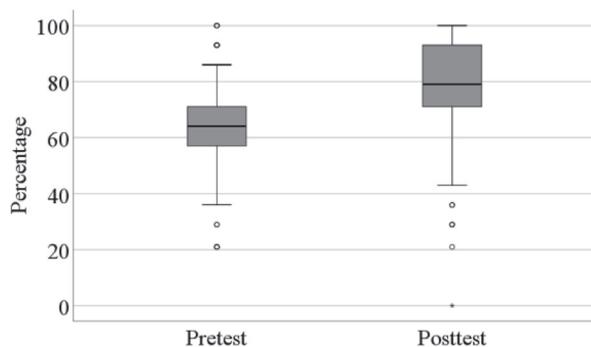


Figure 1. Distribution of *R1* in Pretest and Posttest

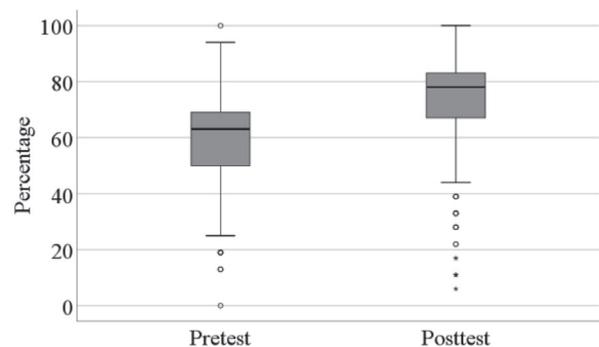


Figure 2. Distribution of *L4* in Pretest and Posttest

5. Conclusion

To conclude, this paper concentrated upon investigating whether there was any improvement in English proficiency demonstrated by the first-year undergraduate students at TPU in Japan, by analysing the results of a standardized English examination of TOEIC Bridge L&R conducted twice, in April 2021 as a pretest and December 2021 as a posttest. The major findings of the present study are summarized as follows.

Concerning the first research question, it may be extrapolated that there was no statistically significant difference between the mean scores of the pretest and posttest.

In addition, with regard to the second research question, it can be inferred, in five out of eight items related to *Abilities Measured* (percentage of correct answers per skill), i.e., *L3* (Short Monologue), *L4* (Listening for Main Idea or Stated Fact), *R1* (Vocabulary), *R3* (Reading for Main Idea or Stated Fact), and *R4* (Short Informational Written Texts), that there might have been some improvements between the pretest and posttest, especially in the aspects of vocabulary and listening for the main idea or stated fact.

It is hoped that some standardized proficiency tests to measure and track English skills and abilities that learners have acquired are appropriately implemented at school on a regular basis, and the results of the examinations are scrutinized closely and properly, in order to employ the eventual outcome effectively in educational settings.

References

- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). <https://doi.org/10.1037/0000165-100>
- Hirai, A. (Writer & Editor). (2012). *Kyoiku shinri-kei kenkyu no tame-no deta bunseki nyumon: Riron to jissen kara manabu SPSS katsuyo-ho* [Introduction to data analysis for educational and psychological research: Learn how to utilize SPSS through theory and practice]. Tokyo Tosho.
- IIBC (n.d.). a. *TOEIC® Bridge Listening and Reading Tests*. The Institute for International Business Communication. Retrieved January 1, 2022, from

<https://www.iibc-global.org/english/toEIC/test/bridge.html>
IIBC (n.d.). b. *TOEIC® Bridge Listening and Reading Tests: Test Format and Content*. The Institute for International Business Communication. Retrieved January 1, 2022, from <https://www.iibc-global.org/english/toEIC/test/bridge/about/format.html>

IIBC (n.d.). c. *TOEIC® Bridge Listening and Reading Tests: Score Range Descriptors/Abilities Measured*. The Institute for International Business Communication. Retrieved January 3, 2022, from https://www.iibc-global.org/english/toEIC/test/bridge/guide04/score_descriptor.html#anchor02

Mizumoto, A. (n.d.). *Koka-ryo keisan shito* [Calculation sheet for effect size] [Microsoft Excel]. <http://www.mizumot.com/stats/effectsizexls>

Yamazaki, D. (2019). Kogaku-bu ichinenji-sei no eigo jitsuryoku shiken kekka ni kansuru bunseki [An analysis of English proficiency testing results collected from first-year undergraduate students at an engineering college]. *Toyama Kenritsu Daigaku Kiyo* [Departmental Bulletin Paper of Toyama Prefectural University], 29, 48-55. https://pu-toyama.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=357&item_no=1&page_id=13&block_id=21

Author Note

Daisuke YAMAZAKI is now at the Centre for Liberal Arts and Sciences, Faculty of Engineering, Toyama Prefectural University.

The author has no conflicts of interest to disclose. This research was supported by a presidential discretionary grant *5-B1* for an already-established educational programme and a fund *1-0* for faculty research in a departmental common budget allocated to the Centre for Liberal Arts and Sciences at Toyama Prefectural University in Academic Year 2021.

The present study owes its completion and accomplishment to the selfless assistance, encouragement and cooperation of a large number of benevolent figures. My special thanks surely go to the students who participated in this study as test subjects, the members of staff at Toyama Prefectural University, and many others concerned. All of the remaining inadequacies are of course mine.

看工学部1年次生の英語実力試験結果に関する縦断的分析

山崎 大介

(工学部教養教育センター)

要約:

この論文は、富山県立大学における学部1年次生を対象に、2021年4月は事前テストとして、2021年12月は事後テストとして、合計2回実施した *TOEIC® Bridge Listening & Reading Tests* の結果を分析することにより、英語力の向上等が見られるのか調査することを最大の目的としている。本研究の目的を達成するため、具体的には、以下の質問に回答する。(1) 事前テストと事後テストにおける平均点には有意な差があるのか。(2) 事前テストと事後テストの間で、英語の能力や技能等における向上が見られるのか。結果として、(1) について、事前テストと事後テストにおける平均点という観点においては、統計的な有意差がないだろうと推定されるかもしれない。そして、(2) に関して、*Abilities Measured* (項目別正答率) における8項目の中の5つにおいて、事前テストと比較して事後テストでは向上が見られる傾向にあるだろうという推定が可能かもしれない。特に、リーディングにおける「語彙」や、リスニングにおける「要点や述べられた事実の理解」という項目において、事前テストと事後テストの間における変化が多少なりともあったのではないかと捉えられるかもしれない。今後、さらに、学習者が習得した英語力を測定する標準化された実力試験が学校等の教育機関で定期的かつ適切に実施され、そうした結果が正しく精査され、教育現場などで効果的に活用されることが期待される。

キーワード: 英語実力試験、学部1年次生、縦断的分析、事前テストと事後テスト、項目別正答率