

# Validation and Linking Scores for the Global Test of English Communication: Analysis Amendment

Prepared for ELS Education Services, Inc.

Minsung Kim, Ph.D.

Tzu-Yun Chin, Ph.D.

Buros Center for Testing, University of Nebraska-Lincoln

September 2019

With questions or comments, please contact: Tzu-Yun (Katherine) Chin tchin@buros.org

### Introduction

The Global Test of English Communication (GTEC) for STUDENTS was introduced in 1998 as a tool intended to assess English communicative abilities of high school students in Japan. Approximately 910,000 students took the GTEC for STUDENTS in 2016. In 2014, a new computer based test (GTEC CBT) was released for use as a university entrance examination. The GTEC CBT is administered all over Japan, including each of the 47 prefectures of Japan. The test scores were accepted at 145 Japanese universities and at eight U.S. institutions for the purpose of college entrance as of 2017.

The GTEC CBT is intended to provide evidence to a university's entrance examination board of a test taker's English ability across four domains: listening, reading, speaking, and writing. Using this approach, the GTEC CBT follows the current teaching trend while also effectively assessing test taker's competency in handling tasks they will likely encounter in real life. GTEC CBT was developed using item response theory. The scale for each section of the GTEC CBT ranges from 0 to 350. The four section scores are summed to create the GTEC CBT total score, ranging from 0 to 1400.

Recently, the Buros Center for Testing (Buros) was contacted by the ELS Educational Services (ELS) to examine the comparability between the GTEC CBT and two widely accepted English proficiency tests for English learners: the TOEFL iBT test and the IELTS Academic test. In this report, we present the validity and linking analyses we conducted and our findings.

#### Data

Buros received the data file from ELS which included 431 examinees with diverse linguistic backgrounds from around the world, including examinees from more than 50 countries. Other demographic variables such as gender, age, and academic background were not provided in the data file. Most examinees took the GTEC CBT in order to meet the English language proficiency requirement for admission to U.S. institutions. These examinees took the GTEC CBT between May 2016 and October 2016. The majority of examinees (94%) also took either the TOEFL iBT or the IELTS Academic within three months prior to or following completion of the GTEC CBT.

The data file included the overall score as well as the section scores for the listening, reading, speaking, and writing of the GTEC CBT for each examinee. In addition, each examinee had scores from either the TOEFL iBT or IELTS Academic included in the data file; similar to the GTEC CBT scores, the total score and the section scores of the listening, reading, speaking, and writing sections from the respective test were recorded separately.

After receiving the data file from ELS, we screened the data for abnormalities such as duplicate or invalid scores. Two examinees were excluded from the sample for further analyses. One student had TOEFL CBT scores rather than TOEFL iBT scores and there was one student in the data file who took the TOEFL iBT twice thus the recording of the examinee's lower score was excluded from analyses. The data cleaning resulted in a sample of 158 examinees who had complete GTEC CBT and TOEFL iBT scores (hereafter referred to as the TOEFL sample) and a sample of 273 examinees who had complete GTEC CBT and IELTS sample). Two students took all three tests and were therefore included in both samples.

#### Analysis

Buros received the data file from ELS which included 431 examinees with the updated GTEC scores. After receiving the data file from ELS, we screened the data for abnormalities such as duplicate or invalid scores. Two examinees were excluded from the sample for further analyses. One student had TOEFL CBT scores rather than TOEFL iBT scores and there was one student in the data file who took the TOEFL iBT twice thus the recording of the examinee's lower score was excluded from analyses. The data cleaning resulted in a sample of 158 examinees who had complete GTEC and TOEFL iBT scores (hereafter referred to as the TOEFL sample) and a sample of 273 examinees who had complete GTEC and IELTS scores (hereafter referred to as the IELTS sample). Two students took all three tests and were therefore included in both samples.

We first reviewed the descriptive statistics and score distributions of the samples. To address **concurrent validity**, we conducted correlational studies of the GTEC with TOEFL iBT and IELTS Academic scores. Pearson product-moment correlation coefficients were calculated between the pair of tests for the total scores and each of the section scores (listening, reading, speaking, and writing). High score correlations are desirable for building interpretable score concordance as the correlations support meaningful comparability between tests (Kolen & Brennan, 2014).

The single-group equipercentile linking method (Kolen & Brennan, 2014) was conducted to build the **concordance tables** between the GTEC and the TOFEL iBT or the IELTS Academic scores. The objective of equipercentile linking is to find the GTEC score estimate that corresponds to a TOFEL iBT or IELTS Academic score at the same percentile rank using the observed score distributions. Equipercentile linking is a successful method that has been used

3

repeatedly for comparing different test scores obtained from a single sample (Sawyer, 2007). This method has been used for various applications including the linking of the SAT scores to the ACT scores (Dorans, 1999) and the linking of the TOEFL iBT scores to IELTS Academic scores (ETS, 2010).

The distributions of the total test and section scores from the TOEFL sample and the IELTS sample were presmoothed using the loglinear model described in von Davier et al (2004). Log-linear presmoothing involves fitting a log-linear model to the sample data to generate a smoother distribution of scores which should more closely mirror the assumed population distribution. For the selection of the number of model parameters to be smoothed, AIC minimization was utilized for estimating test score distributions and for equating, according to the recommendation of Moses and Holland (2009). The AIC minimization process involves fitting multiple loglinear models to the sample data and selecting the model with the lowest AIC value as the best fit for smoothing. Following the AIC minimization process, smoothing was performed to three polynomial degrees (fitting the mean, standard deviation, and skewness of the distributions).

The *equate* package for R (Albano, 2016) was used to apply the linking method described above. Score distributions of each test pair were visually inspected for any irregular equipercentile relationships.

## Results

## **Descriptives and Correlations for the TOEFL Sample**

Similar to the GTEC, the TOEFL iBT also has four sections each for reading, writing, speaking, and listening. The TOEFL iBT section scores range from 0 to 30 and the four section

scores (Reading, Writing, Speaking, and Listening) are summed together to create the TOFEL iBT total score ranging from 0 to 120.

There were 158 students included in the study sample who took both the GTEC and TOEFL iBT tests. Table 1 to Table 5 provide the descriptive statistics (score range, mean, and standard deviation) for GTEC and TOEFL iBT and the correlations between the two tests.

For the TOEFL iBT, the mean of each section was between 15.3 and 17.8, and the standard deviations were between 4.1 and 6.3. On the GTEC scale, the mean scores of the four sections were between 264.4 and 297.3, and the standard deviations were between 31.3 and 48.9. The Pearson correlation coefficients between the GTEC and the TOEFL iBT section scores were around .5, with the highest correlation from the writing section (r = 0.57) and the lowest from the reading section (r = 0.46). The correlation between the GTEC and the TOEFL iBT total scores (r = 0.69) was higher than each of the individual section score correlations.

	TOEFL iBT Reading	GTEC Reading
Possible score range	0-30	0-350
Observed score range	0-29	161-350
Mean	15.7	293.8
SD	6.3	40.7
r	0.4	6

Table 1. Reading Score Statistics and Correlation for the TOEFL Sample (*n*=158)

	TOEFL iBT Writing	GTEC Writing
Possible score range	0-30	0-350
Observed score range	0-27	136-338
Mean	17.1	264.4
SD	5.0	34.9
r	0.5	7

Table 2. Writing Score Statistics and Correlation for the TOEFL Sample (*n*=158)

Table 3. Speaking Score Statistics and Correlation for the TOEFL Sample (*n*=158)

	TOEFL iBT Speaking	GTEC Speaking
Possible score range	0-30	0-350
Observed score range	5-29	169-332
Mean	17.9	266.1
SD	4.1	31.3
r	0.5	5

Table 4. Listening Score Statistics and Correlation for the TOEFL Sample (*n*=158)

	TOEFL iBT Listening	GTEC Listening
Possible score range	0-30	0-350
Observed score range	2-29	20-350
Mean	15.3	297.3
SD	6.2	48.9
r	0.5	56

	TOEFL iBT Total	GTEC Total
Possible score range	0-120	0-1400
Observed score range	24-111	686-1362
Mean	65.9	1121.6
SD	18.2	123.0
r	0.6	59

Table 5. Total Score Statistics and Correlation for the TOEFL Sample (n=158)

## **Descriptives and Correlations for the IELTS Sample**

The IELTS Academic test is also composed of four sections (reading, writing, speaking, and listening). The IELTS Academic section scores ranged from 0 to 9 with .5 increments. The average values of the four IELTS Academic section scores are rounded to the nearest 0.5 in order to create the overall IETLS Academic scores. For example, if the average of the four sections ends in .25, the total score is rounded up to the next half band, and if it ends in .75, the total score is rounded up to the next whole band. Therefore, the IELTS Academic overall score remains on the same scale as its section scores, ranging from 0 to 9.

There were 273 students with both the GTEC and IELTS Academic scores in the data file. Tables 6 to 10 provide information of descriptive statistics (score ranges, means, and standard deviations) and the correlations between the two tests. On the IELTS Academic, the section mean scores ranged from 4.9 to 5.5, and the standard deviations were between 1.0 and 1.1. On the GTEC scale, the section mean scores ranged from 234.8 to 281.7, and the standard deviations were between 41.7 and 46.7. Pearson correlation coefficients for the four sections

ranged from r = 0.48 to r = 0.73. The GTEC total scores were highly correlated with IELTS Academic overall scores (r = 0.82).

	IELTS Academic Reading	GTEC Reading
Possible score range	0-9	0-350
Observed score range	2-9	171-350
Mean	5.0	262.8
SD	1.1	41.7
ľ	0.7	73

Table 6. Reading Score Statistics and Correlation for the IELTS Sample (*n*=273)

Table 7. Writing Score Statistics and Correlation for the IELTS Sample (*n*=273)

	IELTS Academic Writing	GTEC Writing
Possible score range	0-9	0-350
Observed score range	1-8.5	63-350
Mean	4.9	234.8
SD	1.1	46.7
r	0.6	9

Table 8. Speaking Score Statistics and Correlation for the IELTS Sample (*n*=273)

	IELTS Academic Speaking	GTEC Speaking
Possible score range	0-9	0-350
Observed score range	3-9	0-347
Mean	5.5	247.6
SD	1.0	42.4
r	0.4	18

	IELTS Academic Listening	GTEC Listening
Possible score range	0-9	0-350
Observed score range	1-9	153-350
Mean	4.9	281.7
SD	1.1	45.4
r	0.0	66

Table 9. Listening Score Statistics and Correlation for the IELTS Sample (n=273)

Table 10. Total Score Statistics and Correlation for the IELTS Sample (n=273)

	IELTS Academic Overall	GTEC Total
Possible score range	0-9	0-1400
Observed score range	2.5-8.5	602-1385
Mean	5.1	1026.8
SD	1.0	145.5
r	0.8	32

# **Equipercentile Linking and Concordance Tables**

The *equate* package for R (Albano, 2016) was used for equipercentile linking. The score distributions of each test pair were visually inspected for any irregular equipercentile relationships, and no excessive irregularities were found.

The resulting score concordance tables from the smoothed equipercentile linking were presented in Table 11 to Table 15. These concordance tables can be used as a guide for identifying comparable TOEFL iBT or ITELS Academic scores to the GTEC scores. We note that, due to the difference in test specifications and test contents, the score concordance should not be used in the manner that the scores from these different tests are interchangeable. However, the concordance may guide estimation of students' relative standings on a different test.

Results from linking are more precise with larger sample sizes and a greater number of cases within each possible score range. As the TOEFL sample and the IELTS sample had 158 and 273 subjects respectively, there were relatively fewer numbers of subjects in the lower score ranges for each test, causing less accurate results and larger standard errors for low scores. Therefore, more caution is necessary when interpreting the results in the lower end of the score range.

All three tests are shown in each comparison table; however, direct comparison of two test scores between the TOEFL iBT and the IELTS Academic should be avoided, as such a comparison was not intended to be made through the conducted analysis. For this reason, the GTEC score range was placed in the middle of the table to avoid unintended interpretations of the comparison table.

IELTS Academic	GTEC	TOEFL iBT
8-9	346-350	25-30
7	341-345	23-24
6.5	336-340	22
	331-335	21
	326-330	20
6	321-325	19
0	316-320	19
	311-315	10
	306-310	17
	301-305	17
5.5	291-300	16
5.5	281-290	15
	271-280	13-14
	261-270	12
5	251-260	10-11
	241-250	8-9
1.5	231-240	6-7
ч.5	221-230	4-5
1	211-220	2-3
T	201-210	1
3.5	191-200	
3	181-190	
5	171-180	
2.5	161-170	
2.3	151-160	
	141-150	0
	131-140	
2	121-130	
	111-120	
	101-110	
	0-100	

Table 11. Reading Score Conversion Table

Table 12.	Writing	Score	Conversion	Table
-----------	---------	-------	------------	-------

IELTS	GTEC	TOEFL iBT
Academic	246 250	
9	346-350	
	341-345	30
8.5	336-340	20
7.5	331-335	29
	326-330	28
	321-325	27
-	316-320	26
	311-315	25
	306-310	24
6.5	301-305	23
	291-300	21-22
6	281-290	20
Ŭ	271-280	18-19
5.5	261-270	17
	251-260	16
	241-250	14-15
5	231-240	13
	221-230	12
4.5	211-220	10-11
	201-210	9
1	191-200	7-8
4	181-190	5-6
2.5	171-180	4
5.5	161-170	2-3
	151-160	1
3	141-150	
	131-140	
2.5	121-130	
	111-120	0
2	101-110	
1-2	0-100	

IELTS Academic	GTEC	TOEFL iBT
	346-350	
9	341-345	30
	336-340	
	331-335	29
05	326-330	28
8.5	321-325	27
8	316-320	26
0	311-315	25
75	306-310	24
1.5	301-305	23
7	291-300	21-22
6.5	281-290	20
6	271-280	19
0	261-270	17-18
5 5	251-260	16
5.5	241-250	15
5	231-240	14
5	221-230	13
15	211-220	12
4.3	201-210	10-11
4	191-200	9
	181-190	7-8
	171-180	4-6
	161-170	2-3
	151-160	1
3.5	141-150	
	131-140	
	121-130	0
	111-120	U
3	101-110	
1.5-3	0-100	

Table 13. Speaking Score Conversion Table

IELTS Academic	GTEC	TOEFL iBT
7-9	346-350	23-28
6.5	341-345	21-22
(	336-340	19-20
0	331-335	18
	326-330	17
	321-325	16
5.5	316-320	10
	311-315	15
	306-310	15
	301-305	1/
5	291-300	14
	281-290	13
	271-280	12
4.5	261-270	11
	251-260	10
4	241-250	9
	231-240	7-8
	221-230	6
3 5	211-220	5
	201-210	4
3	191-200	•
	181-190	3
2.5	171-180	2
	161-170	
2	151-160	
1.5	141-150	
1	131-140	-
0.5	121-130	
0	111-120	
	101-110	
	0-100	0-2

Table 14. Listening Score Conversion Table

Table 15. Total Score Conversion Table

IELTS Academic	GTEC	TOEFL iBT
9	1376-1400	113-120
	1351-1375	108-112
8-8.5	1326-1350	102-107
7.5	1301-1325	97-101
7	1276-1300	92-96
/	1251-1275	87-91
6.5	1226-1250	82-86
0.3	1201-1225	77-81
6	1176-1200	73-76
0	1151-1175	69-72
	1126-1150	65-68
5.5	1101-1125	61-64
5.5	1076-1100	57-60
	1051-1075	54-56
	1026-1050	51-53
5	1001-1025	48-50
	951-1000	42-47
4.5	901-950	37-41
1	851-900	32-36
4	801-850	28-31
3.5	751-800	23-27
	701-750	18-22
2	651-700	12-17
5	601-650	7-11
2.5	551-600	2-6
	501-550	1
2	451-500	0
	401-450	0
1.5	351-400	0
0-1	0350	0

#### References

- Albano, A.D. (2016). equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software*, 74(8), 1-36.
- Dorans, N. J. (1999). Correspondences between ACT<sup>™</sup> and SAT® I scores. ETS Research Report Series, 1999(1).
- Educational Testing Service. (2010). Linking TOEFL iBT scores to IELTS scores—A research report. Princeton, NJ: ETS.
- Kolen, M. J., Brennan, R. L. (2014). Test equating, scaling, and linking. Methods and practices (Third Edition). New York, NY: Springer.
- Moses, T., & Holland, P. W. (2009). Selection strategies for univariate loglinear smoothing models and their effect on equating function accuracy. Journal of Educational Measurement, 46, 159-176.
- Sawyer, R. (2007). Some further thought on concordance. In N. J. Dorans, M. Pommerich & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 215–230). New York: Springer.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The kernel method of test equating. New York: Springer-Verlag.