



# **Validation and Linking Scores for the Global Test of English Communication: White Paper**

**Prepared for ELS Education Services, Inc.**

**Minsung Kim, Ph.D.**

**Weldon Zane Smith, M.A.**

**Tzu-Yun Chin, Ph.D.**

**Buros Center for Testing, University of Nebraska-Lincoln**

**February, 2017**

With questions or comments, please contact:  
Minsung (Douglas) Kim  
mkim@buros.org  
(402) 472-1414

## **Introduction**

The Global Test of English Communication [GTEC] 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 in 2016. In 2014, a new computer based test 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 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 follows the current teaching trend while also effectively assessing test taker's competency in handling tasks they will likely encounter in real life. GTEC 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 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 in order to meet the English language proficiency requirement for admission to U.S. institutions. These examinees took the GTEC 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.

The data file included the overall score as well as the section scores for the listening, reading, speaking, and writing of the GTEC 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 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 and TOEFL iBT scores (hereafter referred to as the TOEFL sample) and a sample of 273 examinees who had complete GTEC and IELTS Academic scores (hereafter referred to as the IELTS sample). Two students took all three tests and were therefore included in both samples.

## Analysis

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 TOEFL iBT or the IELTS Academic scores. The objective of equipercentile linking is to find the GTEC score estimate that corresponds to a TOEFL 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 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 TOEFL 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 246.2 and 296.6, and the standard deviations were between 41.3 and 52.8. The Pearson correlation coefficients between the GTEC and the TOEFL iBT section scores were around .5, with the highest correlation from the speaking section ( $r = 0.56$ ) and the lowest from

the reading section ( $r = 0.45$ ). The correlation between the GTEC and the TOEFL iBT total scores ( $r = 0.67$ ) was higher than each of the individual section score correlations.

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

	TOEFL iBT Reading	GTEC Reading
Possible score range	0-30	0-350
Observed score range	0-29	161-350
Mean	15.7	293.3
SD	6.3	41.3
<i>r</i>		0.45

Table 2. Writing 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	0-350
Mean	17.1	246.2
SD	5.0	52.8
<i>r</i>		0.53

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	118-339
Mean	17.8	249.3
SD	4.1	42.7
$r$		0.56

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	296.6
SD	6.2	49.5
$r$		0.55

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

	TOEFL iBT Total	GTEC Total
Possible score range	0-120	0-1400
Observed score range	24-111	569-1379
Mean	65.9	1085.3
SD	18.1	147.8
$r$		0.67

### 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 IELTS 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 206.3 to 281.7, and the standard deviations were between 41.7 and 64.8. Pearson correlation coefficients for the four sections ranged from  $r = 0.51$  to  $r = 0.73$ . The GTEC total scores were highly correlated with IELTS Academic overall scores ( $r = 0.82$ ).

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

	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
$r$		0.73



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	0-350
Mean	4.9	206.3
SD	1.1	64.8
$r$		0.70

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	0-350
Mean	5.5	225.0
SD	1.0	55.4
$r$		0.52

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

	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
$r$		0.66

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	457-1400
Mean	5.2	975.7
SD	1.0	171.7
$r$		0.82

### 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 IELTS 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.

Table 11. Reading Score Conversion Table

<b>IELTS Academic</b>	<b>GTEC</b>	<b>TOEFL iBT</b>
8	<b>346-350</b>	27
7	<b>341-345</b>	23
6.5	<b>336-340</b>	22
6	<b>331-335</b>	21
	<b>326-330</b>	20
	<b>321-325</b>	19
	<b>316-320</b>	18
	<b>311-315</b>	
	<b>306-310</b>	17
<b>301-305</b>		
5.5	<b>291-300</b>	16
	<b>281-290</b>	15
	<b>271-280</b>	14
	<b>261-270</b>	13
5	<b>251-260</b>	11
	<b>241-250</b>	9
	<b>231-240</b>	7
4.5	<b>221-230</b>	5
	<b>211-220</b>	3
4	<b>201-210</b>	2
	<b>0-200</b>	0-1

Table 12. Writing Score Conversion Table

Validation and Linking Scores for the Global Test of English Communication

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

Table 13. Speaking Score Conversion Table

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

Table 14. Listening Score Conversion Table

<b>IELTS Academic</b>	<b>GTEC</b>	<b>TOEFL iBT</b>
8	<b>346-350</b>	26
6.5	<b>341-345</b>	21
6	<b>336-340</b>	20
	<b>331-335</b>	19
5.5	<b>326-330</b>	18
	<b>321-325</b>	17
	<b>316-320</b>	16
	<b>311-315</b>	
	<b>306-310</b>	15
<b>301-305</b>		
5	<b>291-300</b>	14
	<b>281-290</b>	13
	<b>271-280</b>	12
	<b>261-270</b>	11
4.5	<b>251-260</b>	10
	<b>241-250</b>	9
	<b>231-240</b>	8
4	<b>221-230</b>	7
	<b>211-220</b>	6
3.5	<b>201-210</b>	5
	<b>191-200</b>	4
3	<b>181-190</b>	3
	<b>171-180</b>	
2.5	<b>161-170</b>	2
	<b>0--160</b>	0-1

Table 15. Total Score Conversion Table

<b>IELTS Academic</b>	<b>GTEC</b>	<b>TOEFL iBT</b>
9	<b>1376-1400</b>	116
	<b>1351-1375</b>	110
8.5	<b>1326-1350</b>	105
8	<b>1301-1325</b>	100
7.5	<b>1276-1300</b>	95
7	<b>1251-1275</b>	91
6.5	<b>1226-1250</b>	87
	<b>1201-1225</b>	82
	<b>1176-1200</b>	78
6	<b>1151-1175</b>	74
	<b>1126-1150</b>	71
	<b>1101-1125</b>	67
5.5	<b>1076-1100</b>	64
	<b>1051-1075</b>	61
	<b>1026-1050</b>	58
	<b>1001-1025</b>	55
5	<b>951-1000</b>	51
	<b>901-950</b>	46
4.5	<b>851-900</b>	41
	<b>801-850</b>	38
	<b>751-800</b>	34
4	<b>701-750</b>	31
	<b>651-700</b>	27
3.5	<b>601-650</b>	24
	<b>551-600</b>	20
3	<b>501-550</b>	17
	<b>451-500</b>	13
	<b>401-450</b>	9
2.5	<b>351-400</b>	5
0-2	<b>0--350</b>	0-2

### Conclusion

This paper assesses some of the validity evidence of the GTEC scores for English communicative abilities. In addition, the GTEC scores were linked to TOEFL iBT scores and IELTS Academic scores to establish concordance. Section scores of the GTEC were moderately correlated to the corresponding TOEFL iBT section scores and more substantially correlated to IELTS Academic section scores. The GTEC total score was highly correlated to the IELTS Academic total score,  $r = 0.82$ , and considerably correlated to the TOEFL iBT total score,  $r = 0.67$ .

According to Cohen's (1988) conventions for interpreting effect sizes, a correlation coefficient of  $r = 0.1$  is thought to represent a weak or small association,  $r = 0.3$  is considered a moderate correlation, and  $r = 0.5$  or larger is regarded as a strong or large correlation. Many reports of educational tests (ACT, 2003; ETS, 2010; NWEA 2011) utilized  $r = 0.5$  as an acceptable value to fulfill a test score comparability. The strength of the correlations between the GTEC total scores and the IELTS Academic and TOEFL iBT total scores were both substantially above .5 thus met the convention standard.

The sample size of the TOEFL sample was smaller than that of the IELTS sample. The correlations of each test score in the TOEFL sample were slightly lower than for the IELTS sample. Generally, smaller samples tend to produce less reliable results, which can lead to lower correlations. In the test score comparisons of the TOEFL sample, only a moderate correlation ( $r = 0.45$ ) was observed for the reading score comparison between GTEC and TOEFL iBT, which may have partially resulted from the relatively smaller number of participants in the TOEFL sample.



In all linking results, large standard errors were observed for low score ranges. A sparse population was observed in the low score range of all tests, and as a result score comparisons within the lower range of the scales are less reliable. Therefore, more caution is required when interpreting the comparison table within the low score ranges.

Last but not least, the concordance table provides general information about scores from different tests. However, the concordance scores resulting from linking, as in the current research, are not as interchangeable or equivalent to one another as equated scores resulting from a more robust non-equivalent anchor test design.

## References

- American College Testing (2003) COMPASS-TABE Concordance Analysis Report. Retrieved Dec 1, 2016, from Retrieved Dec 1, 2016, from <http://www.mccssa.org/wp-content/uploads/COMPASS-TABE-Concordance-Analysis-Report.doc>
- Albano, A.D. (2016). equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software*, 74(8), 1-36.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge
- Dorans, N. J. (1999). Correspondences between ACT™ 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.
- Northwestern Evaluation Association. (2011). Wyoming linking study: A study of the alignment of the NWEA RIT Scale with the Proficiency Tests for Wyoming Students (PAWS). Retrieved Dec 1, 2016, from [https://www.nwea.org/content/uploads/2007/12/WY\\_Linking%20Study.pdf](https://www.nwea.org/content/uploads/2007/12/WY_Linking%20Study.pdf)
- 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.