

Language Research and Development, Inc.

Erwin Tschirner, PhD
445 Hamilton Ave., Ste. 1104
White Plains, NY 10601
USA

June 21, 2016

The ACTFL English Study – Preliminary Report

Technical Report 2016/2-PUB-1

Prepared for:

American Council on the Teaching of Foreign Languages
White Plains, NY
and
Language Testing International
White Plains, NY

Prepared by:

Language Research and Development, Inc.

Dr. Erwin Tschirner
President

The ACTFL English Study – A Preliminary Report

The purpose of the ACTFL English Study is to establish a crosswalk between the Test of English as a Foreign Language (TOEFL) and ACTFL Assessments to help test takers and institutions of higher education (IHE) to better understand the correspondences between TOEFL scores and students' functional proficiency, i.e. their ability to use functional English in real-world situations, both academic and social. The ACTFL Proficiency Guidelines 2012 describe what an individual can do consistently with his or her language ability while listening or reading or in speaking and writing. Because the ACTFL Proficiency Guidelines 2012 provide a developmental perspective, i.e. what a test-taker is able to do now and will be able to do at the next higher level, ACTFL results may be used to determine linguistic areas to be targeted to improve students' proficiency.

The results of the ACTFL English Study may also benefit IHEs by providing a research-based interpretation of how TOEFL scores relate to functional language ability in an English-language context. In addition, IHEs receive more fine-grained information about their students' abilities, including diagnostic feedback to pass on to their students. Moreover, IHEs will learn how ACTFL Assessments might be used to further their own mission with respect to admission and placement and ultimately to professional career goals. Furthermore, they may be able to (re-) evaluate their existing minimal TOEFL scores for admission purposes.

The goal of this study is to administer a total of 400 ACTFL reading and listening assessments, 150 writing assessments, and 150 speaking assessments to foreign students recently admitted to U.S. colleges and universities with known TOEFL scores not older than 9-12 months. Students also complete a background survey to provide further information on their English language background.

Phase 1 of the ACTFL English Study began in July 2015 and was completed in October of the same year. A total of 96 students participated in Phase 1. The ACTFL L&Rcat, a computer-adaptive test consisting of listening and reading assessments, was administered. In addition, students filled out a survey consisting of questions eliciting biographical data. The participating universities provided student TOEFL scores and the date the TOEFL was taken.

The following universities participated in Phase 1 of the study: Cornell University, Georgetown University, Michigan State University, SUNY Plattsburgh, Teachers' College of New Jersey, and Yale University.

42.1% of the students were female, 57.9% were male. 71.6% of the students were graduate students, and 28.4% were undergraduate students. 60% of the students had Chinese as their first language, 11.6% had Portuguese, 8.4% had Hindi, and 4.2% each had Korean and Malay as their first language. Other first languages were Arabic, Bengali, (Indian) English, French, Kirundi,

Norwegian, Spanish, Urdu, and Vietnamese. The average number of years students studied English was 12.47 years (SD = 4.79, Min = 1, Max = 25).

The ACTFL L&Rcat is a computer-adaptive test designed to measure the listening and reading proficiency of students in English. It currently has an item bank consisting of 1,500 items and is based on Rasch measurement. All items were calibrated in 20 separate pilot studies with an overall total of more than 4,000 test-takers to determine difficulty values measured in logits for each individual item. The L&Rcat algorithm selects appropriate items for students on the basis of the correctness of their previous responses and calculates a final person ability value at the end of the test also measured in logits. Person ability values are subsequently rendered as ACTFL sublevels. The following statistics are based on logits and rendered as ACTFL sublevels for ease of comprehension.

The reading assessment took, on average, 34:48 minutes (SD = 8:34; Min = 13:12; Max = 51:34). 4.2% of the students were IM, 13.7% were IH, 9.5% were AL, 37.9% were AM, 29.5% were AH, and 5.3% were S. The mean ACTFL proficiency rating was 7.99 (AM), the lowest rating was NH, the highest one was S. The standard deviation was 2.02. The central 68% of the students, thus, had ratings between AL and AH. The mean TOEFL reading score was 24.77 (SD = 5.44, Min = 12, Max = 30).

The listening assessment took, on average, 30:45 minutes (SD = 5:51; Min = 17:55; Max = 44:20). 4.2% of the students were IM, 13.7% were IH, 9.5% were AL, 37.9% were AM, 29.5% were AH, and 5.3% were S. The mean ACTFL proficiency rating was 7.04 (AM), the lowest rating was IM, the highest one was S. The standard deviation was 1.28. The central 68% of the students, thus, had ratings between AL and AH. The mean TOEFL listening score was 23.93 (SD = 5.84; Min = 9; Max = 30).

The mean TOEFL speaking score was 22.16 (SD = 3.12, Min = 15, Max = 30). The mean TOEFL writing score was 22.22 (SD = 4.85, Min = 10, Max = 30). In Phase 1, no ACTFL speaking or writing assessments were administered.

Reading Proficiency

96 students took the reading assessment. Three assessments were removed because students did not take enough time to read the texts before selecting answers (total times: 2 min, 8 min, and 13 min, respectively).

Figure 1 plots TOEFL reading scores and person ability logits as determined by the ACTFL RPT.

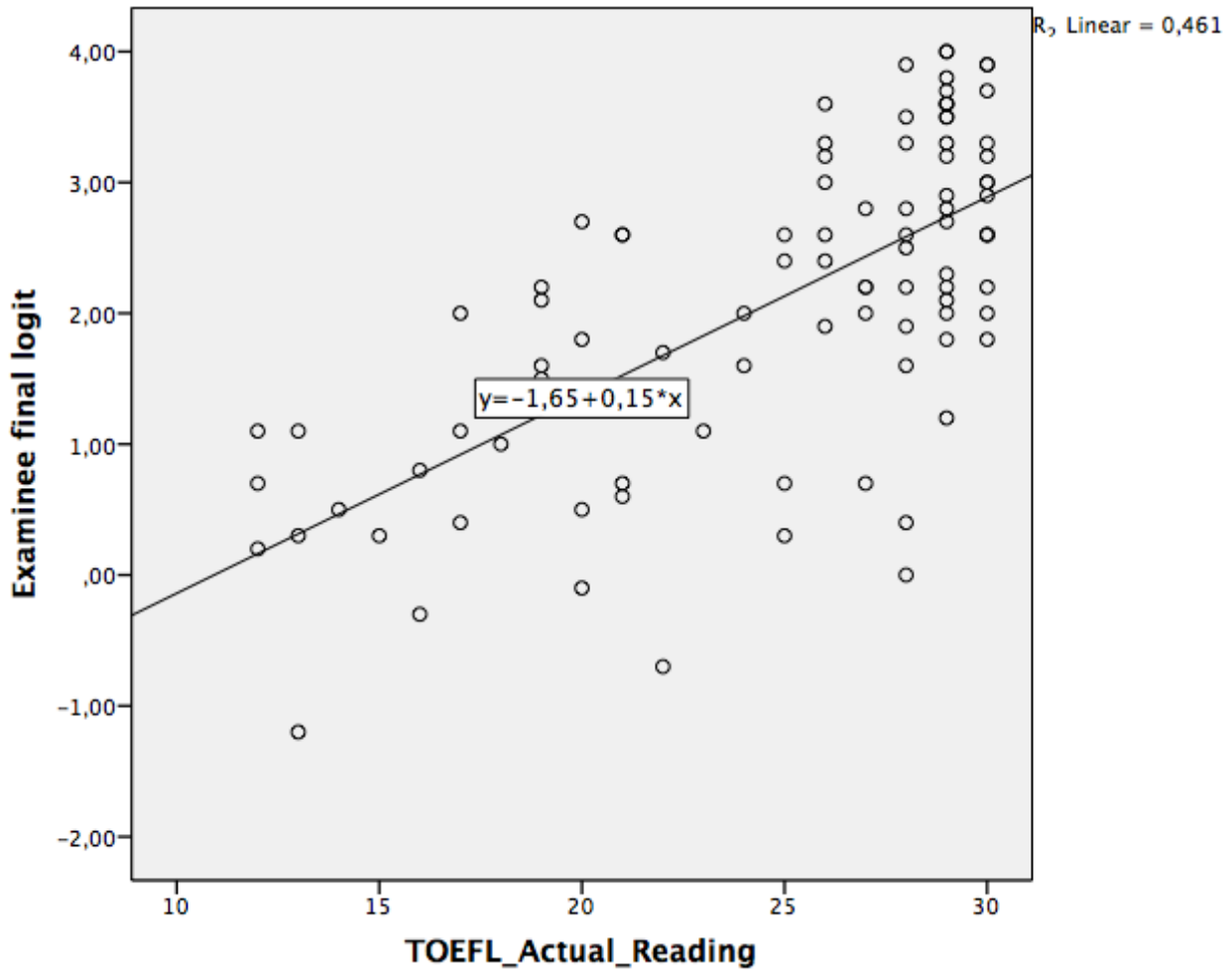


Figure 1. Scatter plot of TOEFL reading scores and ACTFL reading proficiency logits

The correlation between the RPT and the TOEFL reading scores was high: Pearson's $r = .679$, 1-tailed, $p < .001$, $N = 93$. ACTFL proficiency rating accounted for 46.1% of the variance of the TOEFL reading score ($R^2 = .461$). This is a large effect. (Effect sizes above $R^2 = .25$ are considered to be large).

Figure 2 shows a P-P plot of the standardized residuals examining the assumption of normal distribution of the RPT and TOEFL reading data. While there is some evidence of non-normality, the P-P plot by and large shows a linear relationship between RPT person ability logits and TOEFL reading scores.

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: TOEFL_Actual_Reading

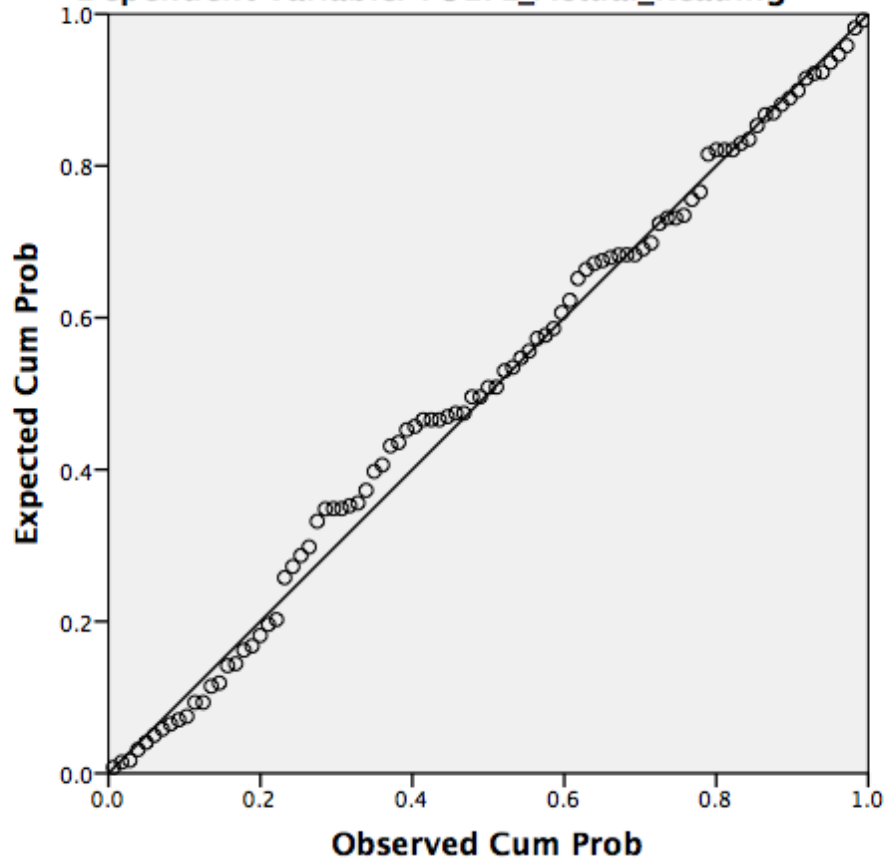


Figure 2. P-P plot of the standardized residuals of the ACTFL reading proficiency logits and TOEFL reading scores

The maximum Cook's Distance was .135 and the maximum Mahalanobis Distance was 7.738 supporting the assumption that there were not outliers.

Table 1 shows the relationship between TOEFL reading scores and ACTFL reading proficiency levels. ACTFL reading proficiency levels were calculated based on their corresponding person ability logit ranges. ACTFL levels are expressed numerically as follows: NL = 1; NM = 2; NH = 3; IL = 4; and so on up to Superior = 10 and Distinguished = 11.

TOEFL_Reading * RPTLPTNumeric Crosstabulation

Count

	RPTLPTNumeric									Total
	3	4	5	6	7	8	9	10	11	
TOEFL_Reading 12	0	0	1	1	1	0	0	0	0	3
13	1	0	1	0	1	0	0	0	0	3
14	0	0	1	0	0	0	0	0	0	1
15	0	0	1	0	0	0	0	0	0	1
16	0	1	0	1	0	0	0	0	0	2
17	0	0	1	0	1	1	0	0	0	3
18	0	0	0	0	1	0	0	0	0	1
19	0	0	0	0	3	2	0	0	0	5
20	0	0	2	0	1	0	1	0	0	4
21	0	0	1	1	0	2	0	0	0	4
22	0	1	0	0	1	0	0	0	0	2
23	0	0	0	0	1	0	0	0	0	1
24	0	0	0	0	1	1	0	0	0	2
25	0	0	1	1	0	2	0	0	0	4
26	0	0	0	0	1	2	0	3	1	7
27	0	0	0	1	0	3	1	0	0	5
28	0	0	2	0	2	3	1	1	2	11
29	0	0	0	0	2	4	3	2	9	20
30	0	0	0	0	1	5	1	4	3	14
Total	1	2	11	5	17	25	7	10	15	93

Table 1. TOEFL Reading Scores and ACTFL Reading Proficiency Crosstabulation

Separate regression analyses were performed to predict TOEFL scores from ACTFL ratings and vice-versa. The mean TOEFL reading score was 24.77 (SD=5.440; N=93), and the mean ACTFL reading proficiency level was 7.99 (SD=2.019), corresponding to Advanced Mid (8.0). Pearson’s correlation between TOEFL reading score and ACTFL proficiency level was .643. Both models explained 41.4% of each other’s results ($R^2=.414$), which is a strong effect.

The linear regression analysis with TOEFL reading score as the dependent variable yielded a significant and large predictive effect of ACTFL Rating on TOEFL score: $p<.001$, Intercept (α): 10.931, Slope (β): 1.733.

The linear regression analysis with ACTFL rating as the dependent variable also yielded a significant and large predictive effect of TOEFL score on ACTFL rating: $p<.001$, Intercept (α): 2.076, Slope (β): .239.

Table 2 shows the minimum TOEFL reading scores predicted by ACTFL reading proficiency levels.

TOEFL	18	20	21	23	25	27	28
ACTFL	IL	IM	IH	AL	AM	AH	S

Table 2. Minimum TOEFL reading scores predicted by ACTFL reading proficiency levels

Table 3 shows the minimum ACTFL reading proficiency levels predicted by TOEFL reading scores.

ACTFL	IM	IH	AL	AM	AH
TOEFL	14	17	21	25	29

Table 3. Minimum ACTFL reading proficiency levels predicted by TOEFL reading scores

The prediction of ACTFL Superior is congruent with a standard-setting study conducted by ETS in 2008, linking TOEFL scores to the Common European Framework of Reference for Languages (CEFR) (Tannenbaum & Wylie, 2008). The reading correspondences in Tannenbaum and Wiley were the following: B1 = a score of 8 on the iBT, B2 = 22, and C1 = 28. Based on the crosswalk between ACTFL and the CEFR (see e.g. Tschirner 2012), the C1 correspondence is the same as the prediction above (C1/S = 28). The B2 correspondence, however, seems to be somewhat off the mark (B2/AM = 25), while the B1/AL correspondence appears to be wildly off the mark. If the preliminary predictions established in the present report hold up, B1 would correspond more appropriately with a TOEFL score between 21 and 23.

Listening Proficiency

96 students took the listening assessment. One student was removed on account of an abnormally low TOEFL listening score of 4, which could have only been due to a malfunction. (Her TOEFL speaking score was 15 and her reading score was 19.)

Figure 3 plots TOEFL listening scores and person ability logits as determined by the ACTFL LPT.

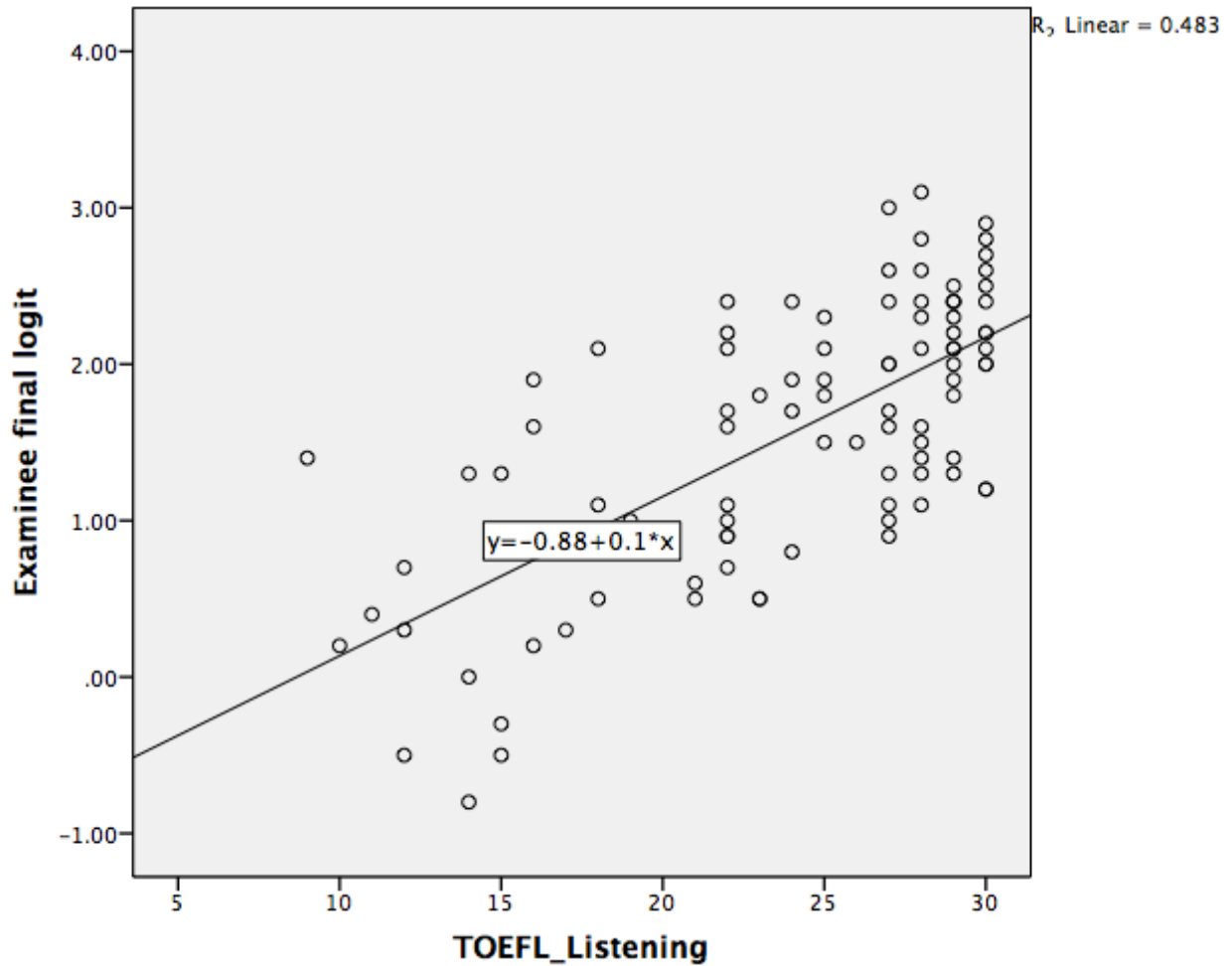


Figure 3. Scatter plot of TOEFL listening scores and ACTFL listening proficiency logits

The correlation between the LPT and the TOEFL Listening Score was high: Pearson's $r = .695$, 1-tailed, $p < .001$, $N = 95$. The ACTFL Proficiency Rating accounted for 48.3% of the variance of the TOEFL Listening Score ($R^2 = .483$). This is a large effect. Effect sizes above $R^2 = .25$ are considered to be large.

Figure 4 shows a P-P plot of the standardized residuals examining the assumption of normal distribution of the RPT and TOEFL Reading data. While there is some evidence of non-normality, the P-P plot by and large shows a linear relationship between RPT person ability logits and TOEFL reading scores.

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: TOEFL_Listening

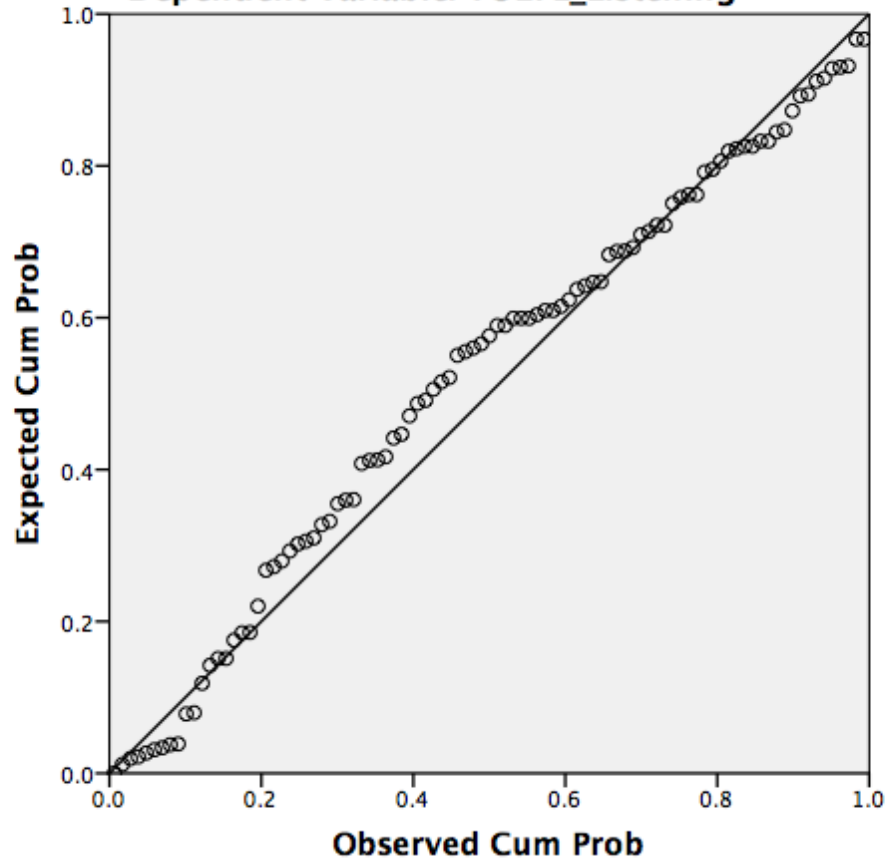


Figure 4. P-P plot of the standardized residuals of the ACTFL listening proficiency logits and TOEFL listening scores

The maximum Cook's Distance was .063 and the maximum Mahalanobis Distance was 7.565 supporting the assumption that there were not outliers.

Table 4 shows the relationship between TOEFL reading scores and ACTFL listening proficiency levels. ACTFL listening proficiency levels were calculated based on their corresponding person ability logit ranges. ACTFL levels are expressed numerically as follows: NL = 1; NM = 2; NH = 3; IL = 4; and so on up to Superior = 10 and Distinguished = 11.

TOEFL_Listening * RPTLPTNumeric Crosstabulation

Count

	RPTLPTNumeric							Total
	4	5	6	7	8	9	10	
TOEFL_Listening 9	0	0	0	1	0	0	0	1
10	0	1	0	0	0	0	0	1
11	0	1	0	0	0	0	0	1
12	1	1	1	0	0	0	0	3
14	1	1	0	1	0	0	0	3
15	2	0	0	1	0	0	0	3
16	0	1	2	2	0	0	0	5
17	0	1	0	0	0	0	0	1
18	0	1	0	1	1	0	0	3
19	0	0	0	1	0	0	0	1
21	0	2	0	0	0	0	0	2
22	0	0	3	4	3	0	0	10
23	0	2	0	1	0	0	0	3
24	0	0	1	2	1	0	0	4
25	0	0	0	3	2	0	0	5
26	0	0	0	1	0	0	0	1
27	0	0	1	5	4	0	1	11
28	0	0	0	5	4	1	1	11
29	0	0	0	4	9	0	0	13
30	0	0	0	2	8	3	0	13
Total	4	11	8	34	32	4	2	95

Table 4. TOEFL Listening Scores and ACTFL Listening Proficiency Crosstabulation

Separate regression analyses were performed to predict TOEFL scores from ACTFL ratings and vice-versa. The mean TOEFL listening score was 23.93 (SD=5.839; N=95), and the mean ACTFL reading proficiency level was 7.04 (SD=1.279; N=95), corresponding to Advanced Low (7.0). Pearson’s correlation between TOEFL listening score and ACTFL proficiency level was .688. Both models explained 47.4% of each other’s results ($R^2=.474$), which is a strong effect.

The linear regression analysis with TOEFL listening score as the dependent variable yielded a significant and large predictive effect of ACTFL Rating on TOEFL score: $p<.001$, Intercept (α): 1.802, Slope (β): 3.142.

The linear regression analysis with ACTFL rating as the dependent variable also yielded a significant and large predictive effect of TOEFL score on ACTFL rating: $p < .001$, Intercept (α): 3.434, Slope (β): .151.

Table 5 shows the minimum TOEFL listening scores predicted by ACTFL listening proficiency levels.

TOEFL	14	18	21	24	27	30
ACTFL	IL	IM	IH	AL	AM	AH

Table 5. Minimum TOEFL reading scores predicted by ACTFL reading proficiency levels

Table 6 shows the minimum ACTFL listening proficiency levels predicted by TOEFL listening scores.

ACTFL	IH	AL	AM	AM
TOEFL	14	21	27	30

Table 6. Minimum ACTFL listening proficiency levels predicted by TOEFL listening scores

The listening correspondences between the CEFR and TOEFL scores established by Tannenbaum and Wylie (2008) were the following: B1 = a score of 13 on the iBT, B2 = 21, and C1 = 26. The ACTFL CEFR crosswalk calls these correspondences into question. If the predictions of this preliminary report hold up, the relationship may be more like the following: B1 = 21; B2 = 27; the highest TOEFL score of 30 would also correspond to B2.

Bibliography

Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. ETS Research Report Series, 2008 (1), i-75.

Tschirner, E. (ed.) (2012). Aligning frameworks of reference in language testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference, Tübingen: Stauffenburg.