



Mapping TOEFL® iBT Scores to the CEFR: An Application of Standard- Setting Methodology

Richard J. Tannenbaum
E. Caroline Wylie

Educational Testing Service

*EALTA Conference June 2007
Sitges, Spain*

Purpose

- Identify scores on TOEFL® iBT corresponding to the six proficiency levels of the CEFR
 - A1 and A2 (Basic)
 - B1 and B2 (Independent)
 - C1 and C2 (Proficient)
- Focus on candidates with “just enough” language skills to be classified into each CEFR level
- Classifications by test section
 - Writing, Speaking, Listening, Reading

Mapping Process

- Expert panel
 - 23 language specialists from 16 EU countries
 - Familiar with TOEFL®, English language instruction, learning and assessment, and the CEFR
- Standard setting approaches
 - Performance-sample (Profile) approach for Writing and Speaking
 - Modified Angoff approach for Reading and Listening

Familiarization/Calibration

- Pre-meeting Assignment – Familiarization with CEFR Levels
 - Review selected tables in the CEFR
 - Write down key skills of candidates just performing at each CEFR level
 - Done for Writing, Speaking, Listening, Reading
- During Meeting – Calibration to CEFR Levels
 - Consensus on skills expected of candidates just performing at each level
 - Pre-meeting assignment, small-group and whole-panel discussions

Sample Level Descriptors

Speaking

B1

- Speaks with some fluency
- Copes with everyday situations
- Briefly gives reasons and explanations
- Describes and briefly explains--with preparation--graphs/tables in field of interest
- Speaks about familiar abstract thoughts, feelings
- Maintains one-on-one conversations, but may need assistance

B2

- Gives clear detailed descriptions and prepared presentations
- Develops clear arguments with relevant examples on wide range of topics in field of interest
- Sustains conversation with degree of fluency and spontaneity
- Takes listener and cultural context into account
- Speaks without causing undue stress to the listener

Profile Approach

- Initial focus on A2, B2, C2 levels
- Review and discuss tasks and rubrics
- Review performance level descriptions (A2, B2, C2)
- Review response profiles across score range
 - Writing 11 profiles
 - Score points 2, 4, 6, 8, 10
 - Speaking 11 profiles
 - Score points 6, 10, 14, 15, 18, 19, 22

Profile Approach

- What score would a “just qualified” A2, B2, C2 candidate earn?
 - Writing: 0 to 10, in half-point increments
 - Speaking: 0 to 24 in one-point increments
- Three rounds of judgments, with feedback and discussion
 - Mean, median, min., max., standard deviation
 - Round 2 includes task-level data--mean scores of candidates in bottom and top quartiles, and overall
 - Round 3 includes percentage of candidates classified A2, B2, C2 based on panel’s recommended cut scores
- Locating the cut scores for A1, B1, C1

Modified Angoff Approach

- What is the probability that a “just qualified” A2, B2, C2 candidate would know the correct answer?
Or
- How many of 100 JQCs would know the correct answer?
- Three rounds of judgments, with feedback and discussion
 - Mean, median, min., max., standard deviation
 - Round 2 includes task-level data—P+ values of candidates in bottom and top quartiles, and overall
 - Round 3 includes percentage of candidates classified A2, B2, C2 based on panel’s recommended cut scores
- Locating the cut scores for A1, B1, C1

Results

Raw Scores and SEJs

	A1	A2	B1	B2	C1	C2
Writing 10 raw pts	-	3 ±.24	5 ±.07	6.5 ±.14	9 ±.10	-
Speaking 24 raw pts	6 ±.14	10 ±.30	15 ±.16	18 ±.31	22 ±.16	-
Listening 34 raw pts	-	-	17 ±.34	26 ±.64	31±.22	-
Reading 45 raw pts	-	-	14 ±.68	29 ±.81	40 ±.55	43 ±.36

Results

Scaled Scores

	A1	A2	B1	B2	C1	C2
Writing 30 scaled pts	-	11	17	21	28	-
Speaking 30 scaled pts	8	13	19	23	28	-
Listening 30 scaled pts	-	-	13	21	26	-
Reading 30 scaled pts	-	-	8	22	28	29

Results

Panelist Evaluations

- All panelists reported that the:
 - pre-meeting assignment was useful preparation
 - instructions and explanations provided were clear
 - training prepared them to complete their standard setting judgments
 - between-round feedback and discussion was helpful
 - standard setting process was easy to follow

Conclusions

- Successfully mapped TOEFL® iBT scores to B1 through C1 levels for all four language skills
- Listening and Reading judged to be too challenging for threshold A-level candidates
- Writing judged to be too challenging for A1 threshold candidates
- Explore convergence with other sources of information



Thank You!

An interim report of this study is available at

<http://www.ets.org//toefl/research.html>

Contact Information
rtannenbaum@ets.org