

**Making use of iBT speaking scores at Carnegie Mellon
Peggy Allen Heidish**

What we did: Correlating iBT speaking to other assessments at the university, Fall 2008 (e.g., ITA test results and ICC placement interviews)

Fall 2008 was the first time a significant number of students at Carnegie Mellon had taken the TOEFL iBT. Before that, only a small percentage of our students had taken iBT and most of those were applicants living or working in an English speaking environment (the iBT was initially introduced in the US, Canada and parts of Europe); earlier scores did *not* represent the normal population of graduate applicants. We used the fall 2008 data to run several studies, including:

- Correlating iBT speaking scores to both ITA tests and initial assessment interviews (required for all incoming NNES, ITAs or not, using our language support)

- Comparing our percentile rankings to the latest percentile rankings by ETS in order to adjust our recommended cut-scores.

Some results:

Correlating iBT Speaking scores to ITA test

- Moderately high correlation for total population (R-sq = 48.7%)
- Standard error for model moderately high, i.e., model predicts “full pass” for iBT ≥ 26 , “provisional pass” possible when accounting for model error
- Implication of above: iBT below approximately 25 may be less predictive of ITA performance
- Caveats - ITA population somewhat small (N = 43) compared to interview population below (N = 276), not heterogeneous (i.e., Indians overrepresented in ITA data)

Correlating iBT Speaking scores to ICC assessment interviews (one-on-one interviews for placement in ICC workshops)

- High correlation for total population (R-Sq = 68.1%)
- High correlation for Taiwan and Korea (R-sq = 77.7%, 69.1%)
- Moderately high correlation for India and China (R-sq = 51.3%, 48.2%), i.e., some Indians tend to score higher while Chinese score both lower and higher than iBT would predict
- Speculate that Indian iBT performance likely influenced by *pronunciation*, Chinese performance by *test preparation*

How we used our data: As a result of our studies, we have a better understanding of how to use the iBT for the following:

1. **Help departments better assess the spoken fluency of applicants** (note: no school of graduate studies, so each department has own standards for admission, own time-line for TA assignments, etc.); our program is often asked to “guide” departments on decision making about language fluency for admission purposes.

- **Offer departments guidelines for recommended cut-scores of applicants.** Scores below are recommended minimum, e.g., lowest level able to handle academic work with on-going ICC support once here; emphasis that departments should set higher cut-scores if students need more than minimal proficiency in any of the skill areas.

Recommended minimum for CM grad admission:

Sub-scores: Reading 21; Listening 21; Speaking 17; Writing 20

Total: 82 (with no one sub-score less than above)

- **Offer rough guideline for understanding scores above recommended minimum:**

IBT Speaking	Suggests following level of spoken fluency
≥ 26	<u>Excellent</u> speaking fluency for all graduate programs
23 -25	<u>Appropriate</u> spoken fluency for most programs
17 - 22	<u>Sufficient</u> spoken fluency, but may be weak for some programs
≤ 16	<u>Insufficient</u> fluency for graduate work

2. **Help departments have realistic expectations about students' possible success on the ITA test** (a performance test that places students into categories defined below):

Rough correlation of iBT speaking to likely success on ITA test

	Provisional pass to TA graduate classes or work one-on-one wt. undergraduate; remedial work required	Provisional pass to TA undergraduate classes; remedial work required	Full pass; can TA all classes; no further language work required
≥ 26	Immediately	Immediately	Within 1 st semester
23 – 25	Immediately	1 - 2 semesters	Can't predict
19 - 22	In 1 st semester	3 - 4 semesters	Can't predict
17 - 18	1 – 2 semesters	4 - 5 semesters	Can't predict

3. **Give our ITA raters additional data to help with testing decisions** (we don't look at the iBT speaking until after we have made a decision, but use both as backup for tricky decisions and to confirm our findings). Helps us feel we are doing the right thing and under scores the validity of our in-house assessment.
4. **Use iBT writing as a way to assess skill set for ITAs/ graders required to do extensive writing.** The ITA test assesses **spoken** language skills and was never intended to provide information about a TA's ability to communicate effectively in written English. Yet an increasing number of departments assign work to ITAs that requires great

skill as a writer: e.g., grading lab reports, journal entries, essays; responding to discussion boards; writing test questions or lab directions, etc. We have recommended that departments use IBT writing, minimum of 24, as screening tool for these ITAs.

Some Observations

- TOEFL prep classes sometimes seem to slightly inflate speaking scores (especially in lower end students, i.e., 17 – 22 speaking)
- Time sensitivity in iBT scores older than one year. We have noticed differences, both higher or lower than expected, given the iBT speaking scores (especially for higher fluency students, i.e., ≥ 23 speaking); seems to depend on what they have been doing with English since the test.
- Indian students – interaction with students in testing sometimes allows fluency to compensate more for accent than is possible on recorded iBT samples.
- Variability and stability of iBT performance - anecdotal stories from students who took the iBT multiple times and received “widely varying” scores on some sections, i.e., the scores did not necessarily show normal incremental improvement arising from more familiarity with the test

Questions for future studies

- Can we predict different rates of progress once here depending on iBT speaking and other background factors (e.g., nationality, motivation, personality, etc.)?
- How accurate will iBT speaking be for applicants who submit 2 year-old scores? We suspect that correlation will be much less accurate because iBT is more sensitive to actual levels of fluency at time of test.