TOEFL iBT™ Update

Eileen Tyson
Director, Client Relations
What English language abilities do your international students need?

• Communicate effectively using all four language skills
  – Read text, listen to lectures, give presentations, write papers

• Communicate at the high level of rigor demanded by graduate study
  – Combine skills to analyze, summarize, synthesize, defend, and argue a point of view

• Teach undergraduates

• Able to fully participate in academic life
Meeting the Needs of Graduate Education

TOEFL® iBT meets those needs

• Emphasis on measuring the ability to *use* English effectively for communication

• Integrated test questions ask students to summarize, evaluate, compare and synthesize information from multiple sources

• Test tasks authentic to graduate study

• 100% academic content

• Backed by extensive and ongoing research to ensure validity
How a test is scored is a critical component of score validity
Important Questions to Ask About Scoring

• Is scoring done by human raters, an automated process, or both?

• If human raters are used, are they calibrated and monitored frequently and consistently for quality control?

• Is rating kept separate and distinct from the test administration process to ensure secure, fair and objective scoring?
The TOEFL® test - unbiased, objective, secure

- Separate the testing administration and scoring process
- Raters see and hear worldwide responses – not just from one country
- Raters do not know the examinee
- Three to six raters anonymously score six Speaking responses
- Four ratings for two Writing responses
Highly trained raters, constantly monitored

• New raters must complete training and pass a certification test

• Must pass calibration test each time they begin scoring or change to a different question type

• Review exemplar responses after a break to refocus attention

• Continually monitored by Scoring Leaders
Scoring You Can Trust

Combining the best of human and automated scoring

July 2009 - TOEFL® program began to use the e-rater® automated scoring engine in a limited way for Writing

- Integrated writing question
  - Two human raters
- Independent writing question
  - One human rater + one automated rating

e-rater® does an excellent job of consistently evaluating mechanical aspects of language
Risks of Using Only Automated Scoring

• Automated scoring cannot measure content and meaning at a significant level

• Questions designed for automated scoring are more coachable and promote rote memorization, not practice of real communication skills

• Questions and resultant scores cannot get at the level of proficiency demanded at the graduate level
The Highest Standards of Scoring

• The TOEFL® iBT integrated tasks require test takers to combine skills and formulate responses at a complex level

• Today no automated scoring can evaluate content and meaning at the sophisticated level tested by TOEFL iBT’s integrated tasks

• Providing you with a more comprehensive picture of a student’s ability to communicate at the graduate level

Scoring that combines

– The judgment of human scoring for content and meaning
– The consistency of automated scoring for mechanical features
New in 2009 - Online Speech Samples

• **Scored** speech response to an actual integrated speaking task from the test

• Extemporaneous response, not prepared in advance

• 60 seconds in length

• Responses available for those testing June 12, 2009 onward

• Register for free service and sign confidentiality statement at [www.ets.org/toefl/speechsample](http://www.ets.org/toefl/speechsample)