



Best Practices in a Digital Age: Artificial Intelligence and Language Assessment

Arum Perwitasari, Ph.D.

Educational Testing Service (ETS) Global

International Association of Teachers of English as a Foreign Language (IATEFL)

IATEFL Webinar, 25 March 2021



What we will cover today

- Artificial Intelligence (AI)
- AI in Language Assessment
 - AI-enabled remote proctoring
 - AI-driven automated scoring
- Resources for teachers
- Q&A



What is AI?

- The simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.
- These machines are able to learn with experience and perform human-like tasks.
 - An intelligent entity created by humans.
 - Capable of thinking and acting rationally and humanely.
 - Capable of performing tasks intelligently without being explicitly instructed.



AI in Language Assessment

- AI has made it possible for language test providers to conduct online assessments with the help of the internet and computer networks.
 - AI-enabled remote proctoring
 - AI-driven automated scoring





AI-enabled Remote Proctoring



AI-enabled Remote Proctoring

- A technology to conduct online assessment that prevents students from possible unfair, fraudulent activities.
- Combines integrated web camera and AI-assisted facial recognition algorithm, and monitoring system.



AI-enabled Remote Proctoring

- Monitors every single test administration from beginning to end, records via video and captures the desktop screen, images and chat logs.
- Tracks nonstandard activities such as a student leaving the room, talking to someone else during the test or leaning away from the web camera.



TOEFL iBT[®] Home Edition Test

- The TOEFL iBT Home Edition test is conducted through artificial intelligence technology and the use of live human remote proctoring serviced by ProctorU[®], the leading proctoring solution for online testing.



TOEFL iBT[®] Home Edition Test:

The role of the proctor

- The at home testing solutions by ETS involve live, remote human proctors to keep an eye on students throughout the test session, in addition to best-in-class AI technology.



TOEFL iBT[®] Home Edition Test:

The role of the proctor

- The proctors confirm the test taker's identity and scan their home environment before testing begins, flag any suspicious activity and intervene if needed.
- Proctors' goal is to detect any wrongdoings during the test session and can cancel the test immediately if there is any attempt to cheat.



TOEFL iBT® Home Edition vs. TOEFL iBT in test centers

- TOEFL iBT Home Edition test is the same TOEFL iBT test taken at a test center:
 - Same content and format
 - Same on-screen experience
 - Same features, like *MyBest*® scores, and instant scoring of Reading and Listening
 - Same price and payment options
 - Same score scales and score reports



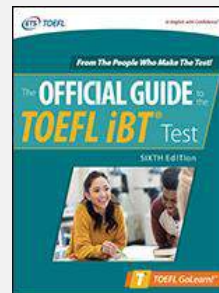
It's the same preparation for the test

- If your students are ready for the TOEFL iBT test, they are ready for the TOEFL iBT Home Edition test too.

Use official TOEFL iBT test prep resources to prepare, including:



TOEFL® Practice Online



The Official Guide to the TOEFL iBT® Test



TOEFL® Test Preparation: The Insider's Guide online course



TOEFL iBT® Free Practice Test



ets.org/toefl/test-takers/ibt/prepare

The **ONLY** differences are ...



Students take
the test from
home



Students use
their own
equipment



Students are
monitored online
by a human
proctor

Equipment Requirements – Computer

To access the test online, students need:



Desktop or laptop, not a tablet or mobile device



PC: Windows[®] operating system, versions 10, 8 or 7
Mac[®]: OS X 10.5 or higher (10.13 High Sierra is recommended)



Chrome[™] or Firefox[®] browser

Equipment Requirements – Computer & Speaker

Students Need:



Internal or
external
microphone



Internal or
external
speaker

Not Allowed:



Students **cannot** use
a headset or
earphones

Equipment Requirements – Camera

Students Need:



A built-in camera in the computer,



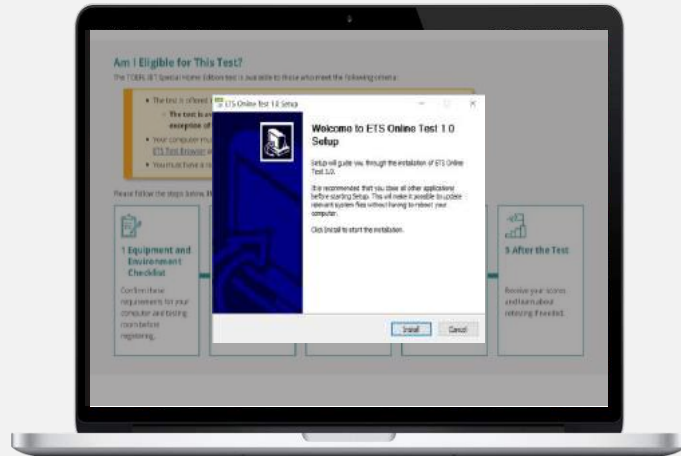
or a separate webcam.



Students will have to show a 360-degree view of the room, including their tabletop surface, before the test

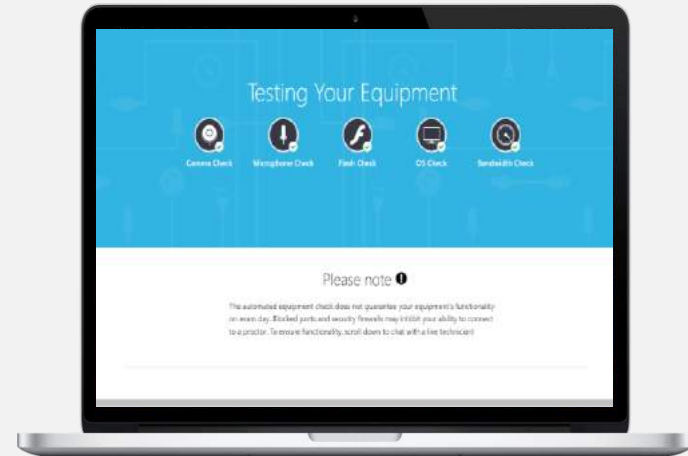
To get their computer ready, students need to ...

Download and install the **ETS Test Browser**



To fully install, students must run the file after downloading it

Run the **ProctorU** equipment check



This checks students' computer, camera, microphone and speaker



AI-driven Automated Scoring



AI-driven Automated Scoring

- AI offers performance-specific feedback, which is not feasible under operational human scoring.
- AI can help score the responses efficiently and reliably, especially for test programs with large test taker volumes.



AI-driven Automated Scoring

- AI cannot fully replace human scoring of spoken and written responses. Human scores can do better some tasks (e.g., evaluate appropriateness of a response), which is why the *TOEFL iBT*[®] test uses the strength of both.

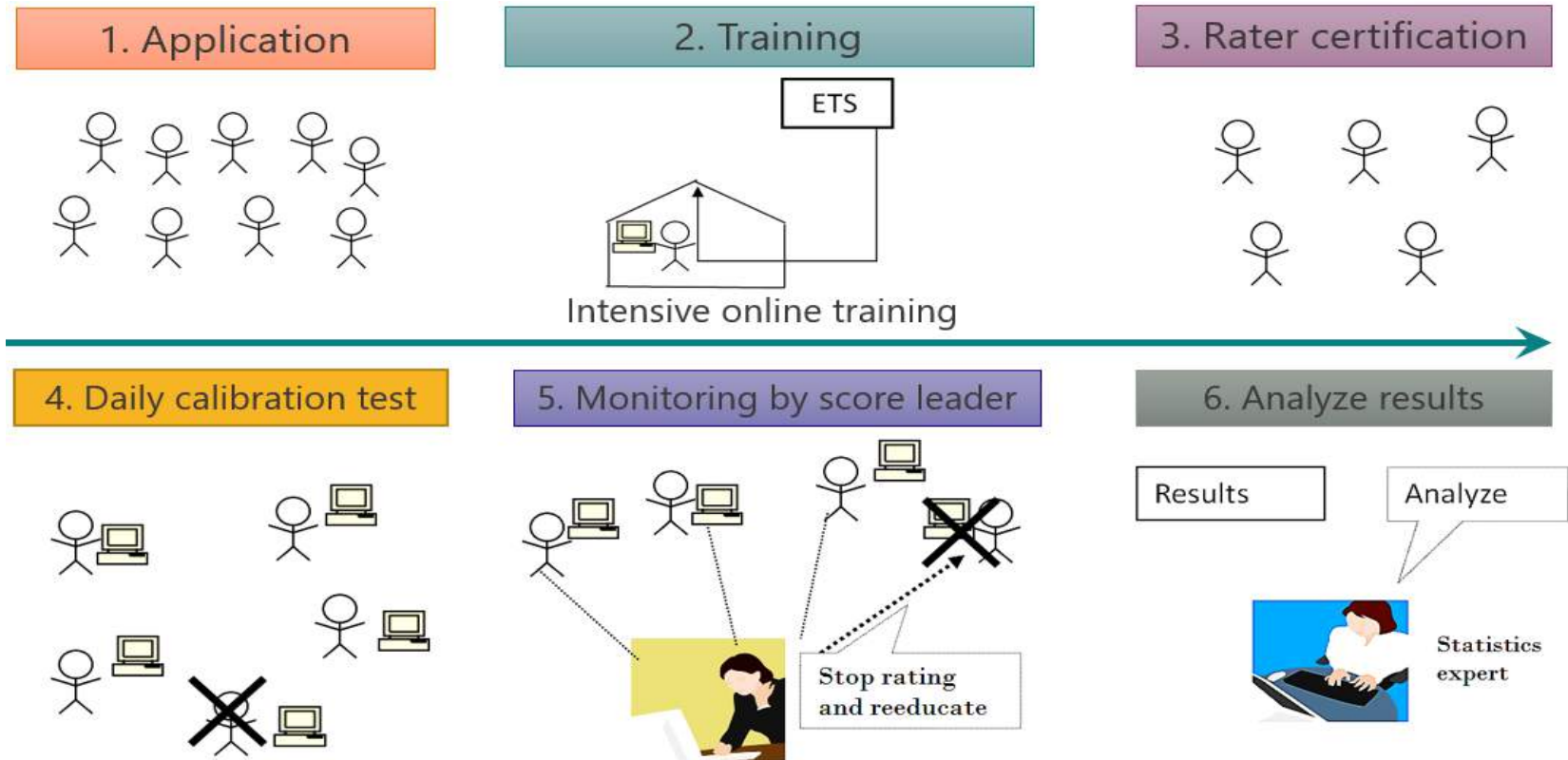


Scoring constructed-responses in the *TOEFL iBT*® test: A hybrid approach

- Scoring for constructed responses in the TOEFL iBT test combines the strengths of machine and human scoring
 - Human raters score all responses
 - Machine scoring is gradually added as automated capabilities mature



Rater training and scoring

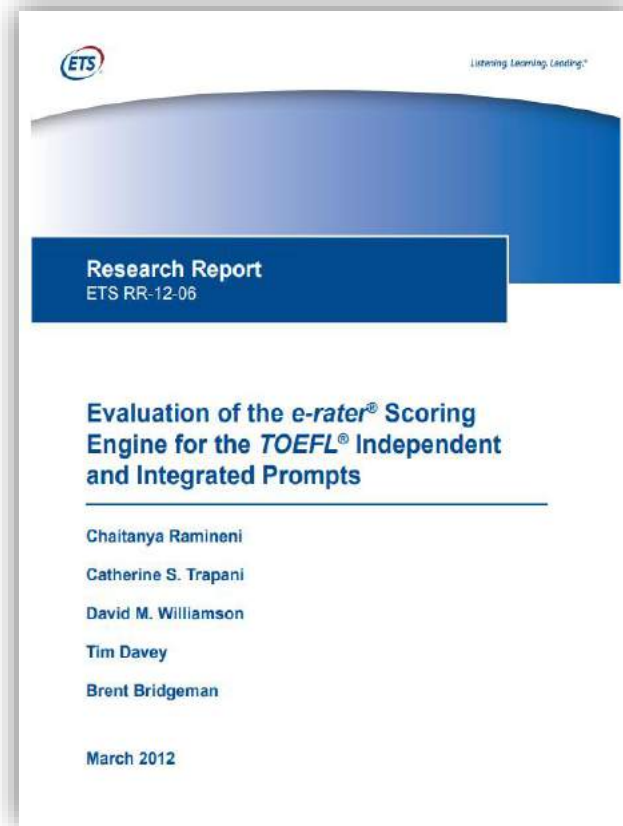
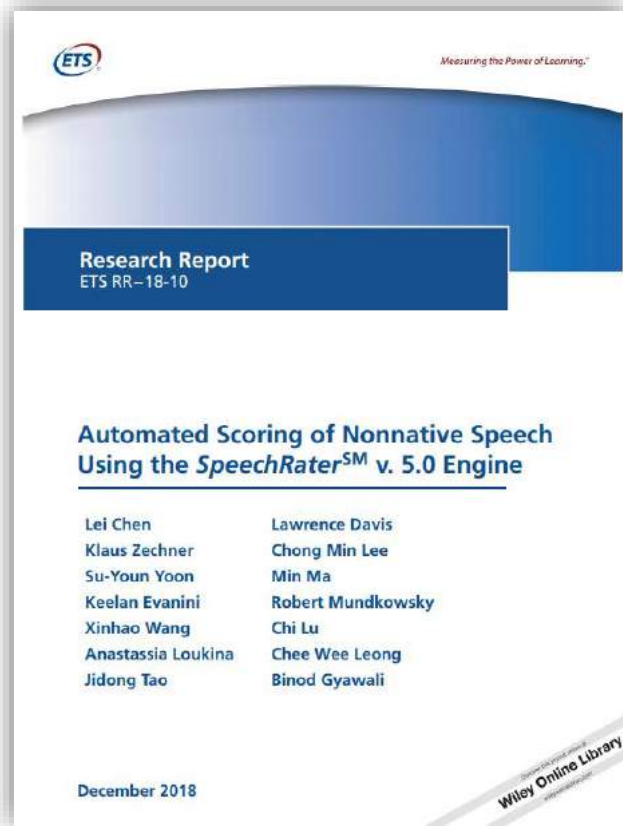


Scoring constructed-responses in the TOEFL iBT® test: A hybrid approach

- The *e-rater*® Scoring Engine has been used in the TOEFL iBT test for almost a decade.
- The *SpeechRater*® Service has started to complement human rating as of August 2019.
- No other major English proficiency test of its kind combines the benefits of AI and human scoring for speaking and writing.



Use of automated scoring technology for constructed-response test sections





**The *SpeechRater*®
Service is used for
assessing speaking**



Assessing speaking in the *TOEFL iBT*[®] test

- 1 independent task and 3 integrated tasks
 - speak about familiar topics drawing on personal experience and background knowledge
 - speak about a topic based on input material on academic course content or campus life
- Total time: 17 minutes (45–60 seconds/response)



Assessing speaking in the *TOEFL iBT*[®] test

- Test takers' responses are recorded and rated through a secure online network on:
 - Topic development
 - Delivery
 - Language use
- The final Speaking test score (0–30) is based on a combination of human and *SpeechRater*[®] scores



Speaking Scoring

- Independent and integrated speaking rubric (holistic)
 - Delivery
 - Language use
 - Topic development
- Ratings are 0–4 from rubrics
 - Converted to a scaled score of 0–30

TOEFL iBT® Test Independent Speaking Rubrics

SCORE	GENERAL DESCRIPTION	DELIVERY	LANGUAGE USE	TOPIC DEVELOPMENT
4	The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse.	Generally well-paced flow (fluid expression). Speech is clear. It may include minor	The response demonstrates effective use of grammar and vocabulary. It exhibits a fairly	Response is sustained and sufficient to the task. It is generally well-developed and coherent; relationships between ideas are clear (or clear progression of ideas).

TOEFL iBT® Test Integrated Speaking Rubrics

SCORE	GENERAL DESCRIPTION	DELIVERY	LANGUAGE USE	TOPIC DEVELOPMENT	
4	The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:	Speech is generally clear, fluid, and sustained. It may include minor lapses or minor difficulties with pronunciation or intonation. Pace may vary at times as the speaker attempts to recall information. Overall intelligibility remains high.	The response demonstrates good control of basic and complex grammatical structures that allow for coherent, efficient (automatic) expression of relevant ideas. Contains generally effective word choice. Though some minor (or systematic) errors or imprecise use may be noticeable, they do not require listener effort (or obscure meaning).	The response presents a clear progression of ideas and conveys the relevant information required by the task. It includes appropriate detail, though it may have minor errors or minor omissions.	Response is mostly coherent and sustained and conveys relevant ideas/information. Overall development is somewhat limited, usually lacks elaboration or specificity. Relationships between ideas may at times not be immediately clear.
3	The response addresses the task appropriately, but may fall short of being fully developed. It is generally intelligible and coherent, with some fluidity of expression, though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:	Speech is generally clear, with some fluidity of expression, but it exhibits minor difficulties with pronunciation, intonation, or pacing and may require some listener effort at times. Overall intelligibility remains good, however.	The response demonstrates fairly automatic and effective use of grammar and vocabulary, and fairly coherent expression of relevant ideas. Response may exhibit some imprecise or inaccurate use of vocabulary or grammatical structures or be somewhat limited in the range of structures used. Such limitations do not seriously interfere with the communication of the message.	The response is sustained and conveys relevant information required by the task. However, it exhibits some incompleteness, inaccuracy, lack of specificity with respect to content, or choppieness in the progression of ideas.	Response is mostly coherent and sustained and conveys relevant ideas/information. Overall development is somewhat limited, usually lacks elaboration or specificity. Relationships between ideas may at times not be immediately clear.
2	The response is connected to the task, though it may be missing some relevant information or contain inaccuracies. It contains some intelligible speech, but at times problems with intelligibility and/or overall coherence may obscure meaning. A response at this level is characterized by at least two of the following:	Speech is clear at times, though it exhibits problems with pronunciation, intonation, or pacing and so may require significant listener effort. Speech may not be sustained at a consistent level throughout. Problems with intelligibility may obscure meaning in places (but not throughout).	The response is limited in the range and control of vocabulary and grammar (some complex structures may be used, but typically contain errors). This results in limited or vague expression of relevant ideas and imprecise or inaccurate connections. Automaticity of expression may only be evident at the physical level.	The response conveys some relevant information but is clearly incomplete or inaccurate. It is incomplete if it omits key ideas, makes vague reference to key ideas, or demonstrates limited development of important information. An inaccurate response demonstrates misunderstanding of key ideas from the stimulus. Typically, ideas expressed may not be well connected or cohesive so that familiarity with the stimulus is necessary to follow what is being discussed.	The response is connected to the task, though the number of ideas presented or the development of ideas is limited. Murky basic ideas are expressed with limited elaboration (details and support). At times relevant substance may be vaguely expressed or repetitive. Connections of ideas may be unclear.
1	The response is very limited in content or coherence or is only minimally connected to the task. Speech may be largely unintelligible. A response at this level is characterized by at least two of the following:	Consistent pronunciation and intonation problems cause considerable listener effort and frequently obscure meaning. Delivery is choppy, fragmented, or telegraphic. Speech contains frequent pauses and hesitations.	Range and control of grammar and vocabulary severely limit (or prevent) expression of ideas and connections among ideas. Some very low level responses may rely on isolated words or short utterances to communicate ideas.	The response fails to provide much relevant content, ideas that are expressed are often inaccurate, limited to vague utterances, or repetitions (including repetition of prompts).	Limited relevant content is expressed. The response generally lacks substance beyond expression of very basic ideas. Speaker may be unable to sustain speech to complete the task and may rely heavily on repetition of the prompt.
0	Speaker makes no attempt to respond OR response is unrelated to the topic.				

Copyright © 2019 by Educational Testing Service. All rights reserved. ETS, the ETS logo, TOEFL and TOEFL iBT are registered trademarks of Educational Testing Service (ETS) in the United States and other countries. ®, ®/TM or ®/TM/CM are trademarks of ETS. 4171019



Independent speaking task

TOEFL Speaking

ETS

Question 2 of 6

VOLUME

Some universities require first-year students to live in dormitories on campus. Others allow students to live off campus. Which policy do you think is better for first-year students and why? Include details and examples in your explanation.



Preparation Time: 15 Seconds
Response Time: 45 Seconds

RESPONSE TIME

00 : 00 : 34



Integrated speaking task (reading input)

TOEFL Speaking  

Question 4 of 6

Read a passage about relict behaviors from a biology textbook. You will have 50 seconds to read the passage. Begin reading now.

Relict Behavior

In general, animals act in ways that help them to survive within their specific habitats. However, sometimes an animal species may display a behavior that no longer serves a clear purpose. The original purpose for the behavior may have disappeared long ago, even thousands of years before. These behaviors, known as relict behaviors, were useful to the animal when the species' habitat was different; but now, because of changed conditions, the behavior no longer serves its original purpose. Left over from an earlier time, the behavior remains as a relict, or remnant, long after the environmental circumstance that influenced its evolution has vanished.



Integrated speaking task (listening input)


TOEFL Speaking

ETS

Question 4 of 6

VOLUME

Now listen to part of a lecture in a biology class.



The image shows a man in a light blue button-down shirt and khaki pants standing at a podium in a classroom. He is looking towards the camera. Behind him is a whiteboard and a wall. The podium has a piece of paper on it.

Integrated speaking task (prompt)

TOEFL Speaking

ETS

Question 4 of 6

VOLUME

Using the example of the pronghorn and lion, explain the concept of a relict behavior.

Preparation Time: 30 Seconds
Response Time: 60 Seconds

PREPARATION TIME

00:00:07



The *SpeechRater*[®] service scoring engine

Adding AI technology to provide the best in measurement

- ETS's *SpeechRater* service uses artificial intelligence (AI) technology to assess and provide feedback on pronunciation, fluency, vocabulary and grammar.

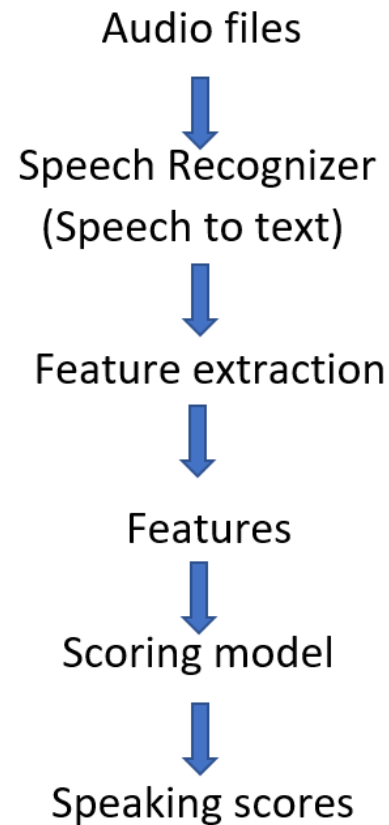
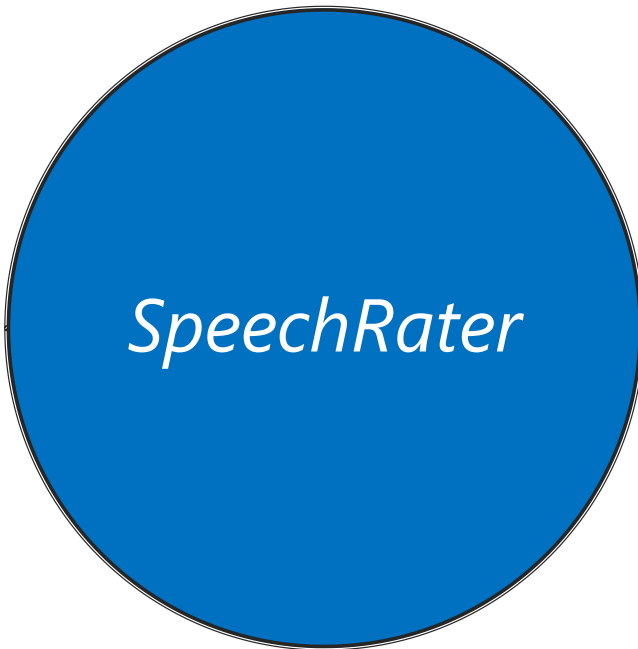


The *SpeechRater*[®] service scoring engine

- The combination of AI and human raters' evaluation of content, meaning and language use provides unmatched accuracy and reliability.
 - Speaking section score now based on more ratings (8) than before (6).
 - 4 human ratings + 4 *SpeechRater* ratings.



The *SpeechRater*[®] service scoring engine



Studies supporting claims about the use of the speaking scores

- Test design and content representativeness
 - test development was informed by reviews of the English-language skills needed for study at English-medium institutions of higher education (Taylor & Angelis, 2008)
 - groups of experts laid out frameworks for a new test design (Butler et al., 2000; Jamieson et al., 2008)
 - teachers' interviews provided support for the content relevance, authenticity and educational appropriateness of integrated test tasks (Cumming et al., 2005)



Studies supporting claims about the use of the speaking scores

- Meaningfulness of test scores
 - responses to the speaking tasks varied predictably, according to proficiency level (Biber & Gray, 2013; Brown et al., 2005)
- Predicting performance in real world
 - scores on speaking tasks have a clear relationship with real-world criterion measures (Brooks & Swain, 2014; Ockey et al., 2015)



Studies supporting claims about the use of the speaking scores

- Usefulness and consequences of test scores
 - speaking scores are useful for the initial screening of ITAs (Xi, 2007)
 - reporting a separate speaking score has prompted greater attention to the development of speaking skills (Wall & Horak, 2005, 2006, 2011)



**The *e-rater*[®]
Automated Scoring
and Feedback Engine
is used for assessing
writing**



Writing Tasks

- 2 Writing Tasks
- 1 Integrated Task
 - Reading/listening/writing
 - Short academic reading and listening material
 - Time: 20 minutes
- 1 Independent Task
 - Response based on personal experience or opinion
 - Time: 30 minutes



Writing Scoring

- Independent writing rubric (holistic)
 - Development of ideas
 - Organization
 - Quality and accuracy of language used
- Integrated writing rubric (holistic)
 - Quality of writing
 - Completeness and accuracy of response
- Ratings are 0–5 from rubrics
 - Converted to a scaled score of 0–30

TOEFL iBT® Test Independent Writing Rubrics

SCORE	TASK DESCRIPTION
5	An essay at this level largely accomplishes all of the following: <ul style="list-style-type: none"> • Effectively addresses the topic and task • Is well organized and well developed, using clearly appropriate explanations, exemplifications and/or details • Displays unity, progression and coherence

TOEFL iBT® Test Integrated Writing Rubrics

SCORE	TASK DESCRIPTION
5	A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.
4	A response at this level is generally good in selecting the important information from the lecture and in coherently and accurately presenting this information in relation to the relevant information in the reading, but it may have minor omission, inaccuracy, vagueness, or imprecision of some content from the lecture or in connection to points made in the reading. A response is also scored at this level if it has more frequent or noticeable minor language errors, as long as such usage and grammatical structures do not result in anything more than an occasional lapse of clarity or in the connection of ideas.
3	A response at this level contains some important information from the lecture and conveys some relevant connection to the reading, but it is marked by one or more of the following: <ul style="list-style-type: none"> • Although the overall response is definitely oriented to the task, it conveys only vague, global, unclear, or somewhat imprecise connection of the points made in the lecture to points made in the reading. • The response may omit one major key point made in the lecture. • Some key points made in the lecture or the reading, or connections between the two, may be incomplete, inaccurate, or imprecise. • Errors of usage and/or grammar may be more frequent or may result in noticeably vague expressions or obscured meanings in conveying ideas and connections.
2	A response at this level contains some relevant information from the lecture, but is marked by significant language difficulties or by significant omission or inaccuracy of important ideas from the lecture or in the connections between the lecture and the reading; a response at this level is marked by one or more of the following: <ul style="list-style-type: none"> • The response significantly misrepresents or completely omits the overall connection between the lecture and the reading. • The response significantly omits or significantly misrepresents important points made in the lecture. • The response contains language errors or expressions that largely obscure connections or meaning at key junctures or that would likely obscure understanding of key ideas for a reader not already familiar with the reading and the lecture.
1	A response at this level is marked by one or more of the following: <ul style="list-style-type: none"> • The response provides little or no meaningful or relevant coherent content from the lecture. • The language level of the response is so low that it is difficult to derive meaning.
0	A response at this level merely copies sentences from the reading, rejects the topic or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank.

...ctic variety, appropriate word choice and

...ly elaborated

...sufficient explanations, exemplifications

...sional redundancy, digression, or

...y and range of vocabulary, though it will

...rm or use of idiomatic language that do

...e, exemplifications and/or details

...is may be occasionally obscured

...Choice that may result in lack of clarity and

...abulary

...Weaknesses:

...to support or illustrate generalizations in

...Following weaknesses:

...is to the task

...topic, or is otherwise not connected to the

...or is blank.



Integrated Writing Task

- Test taker sees a reading (approximately 300 words) for 3 minutes
- Then test taker hears a 2-minute lecture about the same topic from a different perspective or with additional information
- Test taker sees the reading again, then a prompt
- Test taker has 20 minutes to respond to the prompt



Integrated Writing Scoring Guide

- Integrated writing rubric (holistic); descriptors refer to:
 - Quality of writing
 - Completeness and accuracy of response
- Ratings are 0–5 from rubrics
- Raters also work with:
 - Benchmark responses
 - Annotations
 - Key points

TOEFL iBT® Test Integrated Writing Rubrics

SCORE	TASK DESCRIPTION
5	A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.
4	A response at this level is generally good in selecting the important information from the lecture and in coherently and accurately presenting this information in relation to the relevant information in the reading, but it may have minor omission, inaccuracy, vagueness, or imprecision of some content from the lecture or in connection to points made in the reading. A response is also scored at this level if it has more frequent or noticeable minor language errors, as long as such usage and grammatical structures do not result in anything more than an occasional lapse of clarity or in the connection of ideas.
3	A response at this level contains some important information from the lecture and conveys some relevant connection to the reading, but it is marked by one or more of the following: <ul style="list-style-type: none">• Although the overall response is definitely oriented to the task, it conveys only vague, global, unclear, or somewhat imprecise connection of the points made in the lecture to points made in the reading.• The response may omit one major key point made in the lecture.• Some key points made in the lecture or the reading, or connections between the two, may be incomplete, inaccurate, or imprecise.• Errors of usage and/or grammar may be more frequent or may result in noticeably vague expressions or obscured meanings in conveying ideas and connections.
2	A response at this level contains some relevant information from the lectures, but is marked by significant language difficulties or by significant omission or inaccuracy of important ideas from the lecture or in the connections between the lecture and the reading; a response at this level is marked by one or more of the following: <ul style="list-style-type: none">• The response significantly misrepresents or completely omits the overall connection between the lecture and the reading.• The response significantly omits or significantly misrepresents important points made in the lecture.• The response contains language errors or expressions that largely obscure connections or meaning at key junctures or that would likely obscure understanding of key ideas for a reader not already familiar with the reading and the lecture.
1	A response at this level is marked by one or more of the following: <ul style="list-style-type: none">• The response provides little or no meaningful or relevant coherent content from the lecture.• The language level of the response is so low that it is difficult to derive meaning.
0	A response at this level merely copies sentences from the reading, rejects the topic or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank.

T

Copyright © 2019 by Educational Testing Service. All rights reserved. ETS, the ETS logo, TOEFL, and TOEFL iBT are registered trademarks of Educational Testing Service (ETS) in the United States and other countries. IN 00023 (09) WITH CEAF DESIGN IS A trademark of ETS. 10/20/19



In an effort to encourage ecologically sustainable forestry practices, an international organization started issuing certifications to wood companies that meet high ecological standards by conserving resources and recycling materials. Companies that receive this certification can attract customers by advertising their products as “ecocertified.” Around the world, many wood companies have adopted new, ecologically friendly practices in order to receive ecocertification. However, it is unlikely that wood companies in the United States will do the same, for several reasons.

First, American consumers are exposed to so much advertising that they would not value or even pay attention to the ecocertification label. Because so many mediocre products are labeled “new” or “improved,” American consumers do not place much trust in advertising claims in general.

Second, ecocertified wood will be more expensive than uncertified wood because in order to earn ecocertification, a wood company must pay to have its business examined by a certification agency. This additional cost gets passed on to consumers. American consumers tend to be strongly motivated by price, and therefore they are likely to choose cheaper uncertified wood products. Accordingly, American wood companies will prefer to keep their prices low rather than obtain ecocertification.

Third, although some people claim that it always makes good business sense for American companies to keep up with the developments in the rest of the world, this argument is not convincing. Pursuing certification would make sense for American wood companies only if they marketed most of their products abroad. But that is not the case—American wood businesses sell most of their products in the United States, catering to a very large customer base that is satisfied with the merchandise.



Now listen to part of a lecture on the topic you just read about.

Now get ready to answer the question.

You may use your notes
to help you answer.

Directions: You have 20 minutes to plan and write your response. Your response will be judged on the basis of the quality of your writing and on how well your response presents the points in the lecture and their relationship to the reading passage. Typically, an effective response will be 150 to 225 words.

Question: Summarize the points made in the lecture, being sure to explain how they cast doubt on specific points made in the reading passage.

In an effort to encourage ecologically sustainable forestry practices, an international organization started issuing certifications to wood companies that meet high ecological standards by conserving resources and recycling materials. Companies that receive this certification can attract customers by advertising their products as "ecocertified." Around the world, many wood companies have adopted new, ecologically friendly practices in order to receive ecocertification. However, it is unlikely that wood companies in the United States will do the same, for several reasons.

First, American consumers are exposed to so much advertising that they would not value or even pay attention to the ecocertification label. Because so many mediocre products are labeled "new" or "improved," American consumers do not place much trust in advertising claims in general.

Second, ecocertified wood will be more expensive than uncertified wood because in order to earn ecocertification, a wood company must pay to have its business examined by a certification agency. This additional cost gets passed on to consumers. American consumers tend to be strongly motivated by price, and therefore they are likely to choose cheaper uncertified wood products. Accordingly, American wood companies will prefer to keep their prices low rather than obtain ecocertification.

Third, although some people claim that it always makes good business sense for American companies to keep up with the developments in the rest of the world, this argument is not convincing. Pursuing certification would make sense for American wood companies only if they marketed most of their products abroad. But that is not the case—American wood businesses sell most of their products in the United States, catering to a very large customer base that is satisfied with the merchandise.

Cut Paste Undo Redo Hide Word Count 0



Point made in the reading	Counterpoint made in the lecture
<p>Because American consumers have come to distrust frequently-used advertising claims such as 'new' or 'improved,' they won't pay attention to or trust the ecocertified label.</p>	<p>American consumers do pay attention to claims about products when those claims are made by independent consumer agencies.</p>
<p>Since ecocertification adds to the cost of a product, Americans would be unlikely to buy ecocertified products and would choose cheaper, uncertified products.</p>	<p>This is true only if there is a big price difference between two similar products; if an ecocertified product costs only about five percent more, American consumers would accept this in order to buy the product that is better for the environment.</p>
<p>Because American companies sell their products mainly in the U.S., they do not need to compete in the rest of the world where ecocertification is desired by consumers.</p>	<p>American companies must be ready to compete with foreign companies that will soon be selling ecocertified products in the U.S. market.</p>



The *e-rater*[®] scoring engine

- The *TOEFL iBT*[®] test uses the *e-rater* automated scoring engine in a limited and responsible way for the Writing section
 - For each writing task: 1 human rater + 1 automated rating
- Combines the judgment of humans for content and meaning with the consistency of automated scoring for linguistic features



The *e-rater*[®] scoring engine

- In the *e-rater* engine, test taker responses are evaluated in a system where human raters have awarded scores.
- Identifies features associated with writing proficiency in academic English in test taker essays.
- The *e-rater* engine offers a holistic score for a given response to an integrated and independent task besides real-time diagnostic feedback.



The features in the *e-rater*[®] scoring engine

- Content analysis based on vocabulary measures
- Lexical complexity/diction
- Proportion of grammar errors
- Proportion of usage errors
- Proportion of mechanics errors
- Proportion of style comments
- Organization and development scores
- Features rewarding idiomatic phraseology





Combining AI and Human Raters



Important takeaways: Combining AI and raters

- **Human raters** evaluate content, meaning and language in a holistic manner, while **automated scoring** by the *SpeechRater*[®] service and the *e-rater*[®] engine evaluates linguistic features in an analytic manner.



Important takeaways: Combining AI and raters

- The *TOEFL iBT*[®] test is the only large-scale test of academic language proficiency test combining the strengths of automated scoring machines and human scoring, as all spoken and written responses are rated by both multiple human raters and the aforementioned automated scoring engines.



ETS's commitment to validity, reliability and security

- ETS is taking every precaution to ensure that the test students take from home meets the highest standards for validity, reliability and security.
- ETS is and will be committed to developing automated scoring systems to meet these conditions and evaluating test taker responses with the combination of expertise from human raters.





Resources for Teachers



Resources For You

Teachers and Advisors Section

One location for all your resource needs

- *Propell*[®] Workshops
- Teacher Webinars
- Advisor Toolkit

www.ets.org/toefl/teachers_advisors

Videos Available

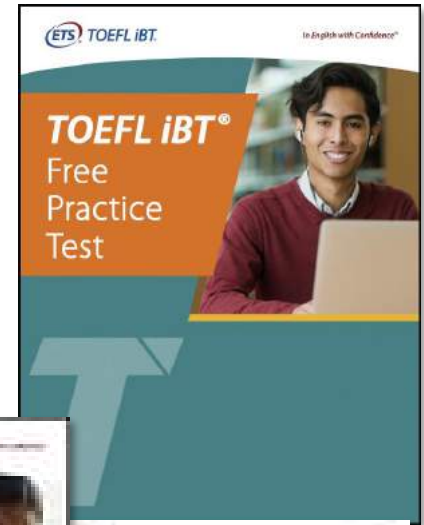
- *TOEFL*[®] Resource Series for Teachers
- Research Behind the *TOEFL*[®] Program
- How ETS Scores the *TOEFL iBT*[®] Test

www.ets.org/toefl/teachers_advisors/video_library



Free *TOEFL*® test prep for your students

- *TOEFL*® Test Preparation: The Insider's Guide: The TOEFL MOOC is a free self-paced course designed by the experts who created the TOEFL test.
- *TOEFL iBT*® Free Practice Test: A full test with all 4 sections and real past test questions, to help them become familiar with the test format and question types.
- *TOEFL iBT*® Practice Sets: Free sets of TOEFL iBT test questions, grouped by test section, in PDF format.
- *TOEFL iBT*® Test Prep Planner: An 8-week preparation plan with tips and activities to build each of the 4 skills.



Inside the *TOEFL*[®] Test

- Video series gives an in-depth look at the Reading, Listening, Speaking and Writing questions, including:
 - Question structure
 - Scoring criteria
 - Sample responses
 - Skill-building tips



Sign up to receive updates:

www.ets.org/toefl/communications

▼ ETS Assessment Program Updates, News & Trends

- College Program Assessments (Student Learning Outcomes)
- GACE® Newsletter for Teacher Certification in Georgia
- GRE® Tests and Services
- GRE® Search Service for Graduate and Professional School Student Recruitment
- HiSET® High School Equivalency Testing Program
- ETS® Professional Educator Programs Communications (*Praxis*® Tests, *PPAT*® Assessment, School Leadership Series, Parapro Assessment and the *ProEthica*® Program)
- TOEFL® Communications
 - Newsletter for Institutions
 - Newsletter for Teachers and Advisors
- TOEFL ITP® Communications
- TOEFL® Search Service for Student Recruitment
- TOEFL® Young Students Series Communications





**Thank you and
stay connected!**

aperwitasari@etsglobal.org

