How Technology is Transforming High Stakes Testing and International Enrollment
Hello! 👋

Jennifer Dewar
Head of Strategic Engagement

Jeffrey Tousignant
Marketing Manager

Geoff Laflair
Assessment Scientist
Agenda

- International Education in 2020
- Duolingo English Test background and overview
- Technology in testing
- Use in practice
- Q&A
Poll!
International Education in 2020
International Enrollment Trends

Int’l Students in the US 1949 - 2019

Percent Change in New Int’l Students

Source: Open Doors® 2019
International Enrollment Trends

Number of Enrolled International Students Drops

"Open Doors" survey finds declines in total numbers of enrolled international students in 2018-19 and continuing declines in new enrollments, though the latter declines are less sharp than in prior two years.

By Elizabeth Redden // November 18, 2019

New International Enrollments at U.S. Colleges

Source: Open Doors

Economics

The International Student Slump Isn't Just About Trump

This worrying development has many causes.

By Justin Fox
August 21, 2019, 7:00 AM EDT

Not as appealing to foreign students? Photographer: Barry Winiker/PhotoLibrary RM

The number of new international students coming to the U.S. for college and university fell in the 2016/2017 and 2017/2018 academic years, according to the Institute of International Education (2018/2019 numbers will be out in November). These marked the first drops in the number of new foreign students since the security crackdown in the aftermath of the terrorist attacks.

International Enrollments Remain Flat, Raising Concerns About America’s Continuing Appeal

By Karin Fischer
November 28, 2019

International enrollment in American colleges and universities stagnated in the fall of 2018, according to the "Open Doors" report released on Monday by the Institute of International Education and the U.S. State Department. The 0.05-percent increase in total enrollment is the smallest in the 70 years the data have been tracked, except for a couple of years of outright declines following the attacks of September 11, 2001.

The number of new international students fell for the third year in a row, although that 1-percent decline was not as steep as the nearly 7-percent drop the previous year.

Fewer New Students From Abroad

The number of new international students arriving on American campuses fell for the third year in a row. New international enrollments have tumbled more than 10 percent since the fall of 2015.
Testing in the News

The Next Testing Scandal

Don’t dwell on the Varsity Blues scandal, writes Ben Paris. Focus on the ways there are advantages, in plain sight, to cheat.

By Ben Paris // January 21, 2020

Almost one year ago, a college admissions cheating conspiracy made headlines, partially because the behavior involved was memorably egregious, and partially because some of the participants were celebrities. It was a great story, but it might have given the impression that this kind of wholesale fraud is more common than it is.

U. of California Faces Bias Lawsuit Over ACT/SAT Requirement

By Eric Hoover | OCTOBER 29, 2019

Updated (10/29/2019, 6:43 p.m.) with comment from the College Board.

Lawyers representing students, the Compton Unified School District, civil-rights groups, and college-access organizations said on Tuesday that they planned to sue the University of California unless it drops its ACT/SAT requirement.

Education organizations cancel IELTS, GRE, TOEFL tests across China as virus spreads

SHANGHAI (Reuters) - Educational testing organizations have canceled the IELTS, GRE, and TOEFL tests across China for the remainder of January and all of February, according to social media posts published on official accounts on Monday.

Reporting by Josh Horwitz and Winni Zhou; Editing by Tom Hogue

Our Standards:  The Thomson Reuters Trust Principles.
Accessibility for International Applicants

NACAC AND INTERNATIONAL ACAC URGE FLEXIBILITY FOR APPLICANTS TESTING OUTSIDE THE US

Media Contact:
Lindsay Addington
703-299-6829

Arlington, VA (Oct. 23, 2019) - The College Board recently announced significant score cancellations from the Oct. 5, 2019 administrations in Egypt and Hong Kong due to issues related to test security. As a first step NACAC and International ACAC remind colleges and universities that these disruptions could significantly affect applications to your institutions for students who tested in these countries.

Short-Term Considerations

Considering upcoming application deadlines, NACAC and International ACAC urge colleges and universities that require or recommend standardized testing for admission or scholarships to consider the following points.

- The delay in receiving scores, or score cancellations, from the Oct. 5 SAT may disrupt the college preparations of affected students, causing additional stress during an already demanding process.

- The next international SAT administration takes place in December. There is no access to a November SAT date internationally.

- Affected students are limited in their ability to re-take the test, or to travel to other locations to take the test, posing significant equity concerns.

What Colleges and Universities Can Do:
Purchasing Power Parity

$160  $15,333
$614
$1,351
$11,691
Test Centers vs Internet Access

1 Test Center per 1,926,260 people

1 Internet user per 2.2 people
Test Centers in Brazil

- **Amazona**
  - 4M people
  - 1 city, 6 dates

- **Rondônia**
  - 1.5M people
  - Nearest test center 14+ hours away

- **Bahia**
  - 15.2M people
  - 1 city, 8 dates

*July 2018*
Duolingo Background

Duolingo’s mission is to make education accessible to all!

Language learning app

- 30+ distinct languages
- iTunes + Google Play “App of the Year”
- Most downloaded education app
- Over 300 million users
Behind the Mission

Luis von Ahn Ph.D., Duolingo’s co-founder and CEO

- Born and raised in Guatemala
- Inventor of CAPTCHA, reCAPTCHA
- Former CS professor at CMU
- MacArthur Fellow
- Lemelson-MIT Recipient
# English Proficiency Options

<table>
<thead>
<tr>
<th></th>
<th>TOEFL</th>
<th>IELTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>$200+</td>
<td>$215+</td>
</tr>
<tr>
<td><strong>Locations</strong></td>
<td>4500</td>
<td>1100</td>
</tr>
<tr>
<td><strong>Test Length</strong></td>
<td>3 Hours</td>
<td>3 Hours</td>
</tr>
<tr>
<td><strong>Turnaround</strong></td>
<td>14 Days</td>
<td>13 Days</td>
</tr>
<tr>
<td><strong>Sharing</strong></td>
<td>5 Included</td>
<td>4 Included</td>
</tr>
</tbody>
</table>
Test Overview
Accessible

No test centers. No appointments. Accessible online, on demand for only $49 with unlimited sharing, and completely free for schools.
Comprehensive

Measures speaking, writing, reading, and listening. Integrates a proficiency score, video interview, and writing sample.
Fast

Take a test in under 60 minutes, and students receive results in under 48 hours.
Secure

Rigorous security protocols, individual test proctoring, and computer vision + biometrics help prevent and detect rule breaking.
Accurate

Reliability measures and correlation with other tests support the use of test for university admissions.
<table>
<thead>
<tr>
<th></th>
<th>TOEFL</th>
<th>IELTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>$200</td>
<td>$215</td>
</tr>
<tr>
<td>Locations</td>
<td>4500</td>
<td>1100</td>
</tr>
<tr>
<td>Test Length</td>
<td>4 Hours</td>
<td>3 Hours</td>
</tr>
<tr>
<td>Turnaround</td>
<td>14 Days</td>
<td>13 Days</td>
</tr>
<tr>
<td>Sharing</td>
<td>5 Included</td>
<td>4 Included</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$49</td>
<td>Unlimited</td>
</tr>
<tr>
<td></td>
<td>&lt; 1 Hour</td>
<td>2 Days</td>
</tr>
<tr>
<td></td>
<td>Unlimited</td>
<td></td>
</tr>
</tbody>
</table>
Technology in Testing
Technology in Testing

- Developing the test items
- Administering the test
- Scoring the test
- Securing the test
Item Development

CEFR LEVELS

BEGINNER
A1   A2

INTERMEDIATE
B1   B2

ADVANCED
C1   C2
Item Development

1. Expert-annotated CEFR texts & wordlists
2. Natural English language sources
3. Statistical models for CEFR difficulty estimation
4. Item-generation techniques
5. Large pool of CEFR-aligned items for CAT administration
DEVELOPING

Relationship with CEFR

(5-fold cross-validation; wordlists compiled from Capel, 2010/2012; North et al., 2011; other sources)
DEVELOPING Peer-Reviewed Publication

Transactions in Applied Computational Linguistics (Forthcoming)
February 2020

Machine Learning Driven Language Assessment

Abstract

Language proficiency testing is an increasingly important part of global society. The need is growing for automated language assessment that can provide both high-quality performance feedback and meaningful scores. In this paper, we present a novel approach to language assessment that combines machine learning with human expertise for feedback and scoring.

1. Introduction

Language proficiency testing is an increasingly important part of global society. The need is growing for automated language assessment that can provide both high-quality performance feedback and meaningful scores. In this paper, we present a novel approach to language assessment that combines machine learning with human expertise for feedback and scoring.

2. Background

This section provides a brief overview of existing approaches to language assessment.

3. Methodology

Our approach is based on a combination of machine learning and human expertise. We use a combination of automated and human-generated feedback to improve the accuracy and fairness of the assessment.

4. Results

Our results show that our approach is effective in providing accurate and fair feedback. We also show that our approach is scalable and can be applied to other languages.

5. Conclusion

In conclusion, our approach to language assessment combines the strengths of machine learning and human expertise to provide accurate and fair feedback. We believe that this approach has the potential to improve language assessment for all learners.

Figure 1: Sample output of our machine learning-driven feedback system.

Table 1: Comparison of scores generated by human and automated systems.

Table 2: Performance of our system on a variety of language proficiency tests.

Figure 2: Graph showing the improvement in language proficiency scores over time.

Figure 3: Heat map showing the distribution of scores across different proficiency levels.
Test Administration

- Computer adaptive test ensures virtually unique test experience
- Draws from an extremely large question pool
  - Questions generated by language assessment experts and machine learning
- Monitor test item exposure and retire questions as necessary
Computer Adaptive Testing

Note: item types are equally represented in the test, to represent all skills
Each formant is shown 3–7 times (6 on average)
% tests that use a given item (Way, 1998)

Most high-stakes CATs are significantly higher than ours (e.g., GRE is 14%)

0.10% (~1 in 1000 tests)
Security: exam overlap rate

- % items shared between any two test administrations (Chen et. al, 2003)

- We haven’t found any other high-stakes CATs that disclose their overlap rates publicly

0.43%  
(< 1 in 100 items shared)
Test Item Security

1000+ Number of tests a student would need to take in order to see a single repeated question

~41 Years Time for a test taker to see a single repeated question ...by then, it would be long retired
Scoring the Test

LANGUAGE FEATURES

- Grammatical accuracy
- Grammatical complexity
- Lexical sophistication
- Lexical diversity
- Task relevance
- Length | rate | acoustic

CONNECTION: SPEAKING/WRITING

- Content coverage
- Sophistication
- Accuracy | rate | acoustic
Extended Writing Prompts

Cohen’s Kappa

- Human ↔ Machine agreement: 0.79
- Human ↔ Human agreement: 0.77

Pearson’s $r$

- Reliability: 0.81

$n = 3,626$
Extended Speaking Prompts

Cohen’s Kappa

- Human ↔ Machine agreement: 0.82
- Human ↔ Human agreement: 0.68

Pearson’s $r$

- Reliability: 0.80

$n = 3,966$
## Relationship with other tests

### Correlation

<table>
<thead>
<tr>
<th>Relationship</th>
<th>r</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>*TOEFL ↔ IELTS</td>
<td>.73</td>
<td>1,153</td>
</tr>
<tr>
<td>DET ↔ IELTS</td>
<td>991</td>
<td></td>
</tr>
</tbody>
</table>

*ETS, 2010*
### Relationship with other tests

<table>
<thead>
<tr>
<th>Relationship</th>
<th>r</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>*TOEFL ↔ IELTS</td>
<td>.73</td>
<td>1,153</td>
</tr>
<tr>
<td>DET ↔ IELTS</td>
<td></td>
<td>991</td>
</tr>
<tr>
<td>DET ↔ TOEFL</td>
<td>.77</td>
<td>2,319</td>
</tr>
</tbody>
</table>

*ETS, 2010

**Correlation**

![Scatter plot showing correlation between DET and TOEFL scores](image)

**Table:**

<table>
<thead>
<tr>
<th>Relationship</th>
<th>r</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>*TOEFL ↔ IELTS</td>
<td>.73</td>
<td>1,153</td>
</tr>
<tr>
<td>DET ↔ IELTS</td>
<td></td>
<td>991</td>
</tr>
<tr>
<td>DET ↔ TOEFL</td>
<td>.77</td>
<td>2,319</td>
</tr>
</tbody>
</table>

**Diagram:**

Scatter plot with a trend line indicating the correlation between DET and TOEFL scores.
Certification Process

We record every test experience via:
- Webcam
- Microphone
- Screen
- Keyboard
- Mouse

Machine learning, biometrics authentication, and computer vision analyze the test for irregular response patterns and suspicious behavior to help identify rule breaking and generate a “risk index”

Independent, anonymous proctors are assigned to the test session and are trained to look for 75 different behaviors such as eye movement and background noise

A second independent, anonymous proctor reviews the same session for over 75 different behaviors. Tier 2 is comprised of professional, certified ESL or English language educators

As needed, an additional round of review is added for tests where there is disagreement between the earlier rounds or further examination is needed

Based on this process, the results are certified or not certified
We use **biometrics and computer vision** to help verify examinee identity and ensure test integrity
- Face detection & ID verification
- Voice recognition
- Keystroke/mouse biometrics
- Eye tracking, lip syncing
Proctor: Test Taker Ratio

1:25

- Varies depending on the test and center, but on average, 1 proctor is supervises over 25 test takers ✗
- Proctors are often contracted locally so they are vulnerable to bias or conflict of interest ✗

2:1

- Since there are multiple independent rounds of proctoring, each test session is individually reviewed by two, and sometimes three, different proctors ✔
- Proctors are anonymous and have no interaction with test takers ✔
In Practice
Test Countries

Afghanistan Albania Algeria American Samoa Angola Antigua and Barbuda Argentina Armenia Australia Austria Azerbaijan Bahamas Bahrain Bangladesh Barbados Belarus Belgium Belize Benin Bhutan Bolivia Bosnia and Herzegovina Botswana Brazil Bulgaria Burkina Faso Burundi Cabo Verde Cambodia Cameroon Canada Central African Republic Chile China Colombia Congo Costa Rica Côte d’Ivoire Croatia Cuba Cyprus Czechia Denmark Dominican Republic Ecuador Egypt El Salvador Eritrea Estonia Fiji Finland France Gabon Gambia Georgia Germany Greece Guatemala Guinea Guinea-Bissau Honduras Hong Kong Hungary Iceland India Indonesia Iran Iraq Ireland Israel Italy Jamaica Japan Jordan Kazakhstan Kenya Kuwait Kyrgyzstan Laos Latin America Lebanon Lesotho Liberia Libya Lithuania Luxembourg Macao Macedonia Madagascar Malawi Malaysia Mali Malta Mauritania Mauritius Mexico Mongolia Netherlands New Zealand Nicaragua Niger Nigeria Norway Oman Pakistan Palestine Panama Papua New Guinea Paraguay Peru Philippines Poland Portugal Puerto Rico Qatar Republic of Korea Republic of Moldova Romania Russia Rwanda Saint Lucia Saudi Arabia Senegal Serbia Sierra Leone Singapore Sint Maarten Slovakia Slovenia Somalia South Africa South Sudan Spain Sri Lanka Sudan Suriname Sweden Switzerland Syria Taiwan Tajikistan Tanzania Thailand Togo Trinidad and Tobago Tunisia Turkey Turkmenistan Uganda Ukraine United Arab Emirates United Kingdom United States of America Uruguay Uzbekistan Venezuela Vietnam Virgin Islands Yemen Zambia Zimbabwe
Accepting Institutions

Appalachian State  Babson College  Bard College  Barnard College  Baylor University  Boise State University  Bowdoin College  Clark University  Clarkson University  Colby College  Columbia University  Creighton University  Dallas Baptist University  Dartmouth College  Duke University  Elon University  Emerson College  Emory University  Fairfield University  Florida Tech  Furman University  George Washington  Georgetown University  Georgia Tech  Goldey-Beacom College  Hong Kong University  Houston Community College  Iowa State University  Ithaca College  Johns Hopkins University  Jacobs University  Lafayette College  Lehigh University  Lindenwood University  Loyola Marymount University  Macalester College  Marist College  Miami Dade College  Middle Tennessee State  Middlebury College  Milwaukee School of Engineering  Montana State University  The New School  Northeastern University  Northern Arizona University  Occidental College  Orange Coast College  Pasadena City College  Pepperdine University  Rhode Island School of Design  Rhodes College  Robert Morris University  Rollins College  Santa Clara University  Sewanee  Southeast Missouri State  Stanford University  Stetson University  Suffolk University  Swarthmore College  Tennessee Tech  Texas Christian University  Texas Tech University  The New School  The University of Chicago  Trinity University  Troy University  Tufts University  Union College  University of Akron  University of Alabama  U Alabama - Birmingham  UC Los Angeles  UC Santa Barbara  University of Connecticut  University of Evansville  University of Kentucky  University of La Verne  University of Missouri - KC  University of New Hampshire  University of North Texas  University of Notre Dame  University of Oregon  University of Pittsburgh  University of Rhode Island  University of Southern California  University of Tampa  University of Tulsa  University of Vermont  University of Wyoming  Valparaiso University  Vanderbilt University  Washington and Lee University  WashU  Wellesley College  Wesleyan University  Whitman College  Widener University  Williams College  Yale University

500+ More at englishtest.duolingo.com/institutions
Admissions Use Cases

- Completing apps, borderline candidates
- Reaching more markets
- Scholarship demographics
- Email campaigns
- Triage for cancelled scores
- Athletics
Access Program

Supporting low-income students

We're waiving test fees for organizations who work with high-achieving, low-income students.

APPLY NOW

englishtest.duolingo.com/access
Thanks!
Questions?

Jennifer Dewar  |  jdewar@duolingo.com
Jeff Tousignant  |  jt@duolingo.com
Geoff Laflair  |  geoff@duolingo.com