

Evidence Based Criteria and Methods:

Role of Standardized Tests in Admissions -Using the *TOEFL*[®] and *GRE*[®] Tests for Graduate Admissions in Europe

Zeynep Kaptan

Academic and Government Relations Coordinator - ETS Global

EURIE Eurasia Higher Education Summit 2017, Turkey
March 23, 2017



What We Will Discuss Today

- The definition of standardized testing and its value for master level admissions
- The *TOEFL*® testing program: assessing academic English skills
- The *GRE*® testing program: assessing master level skills
- Best practices in score and ETS data in recruitment and admissions

The Definition of Standardized Testing

What is a “Standardized” Test?

- Standardized test means:
 - Same or comparable tasks
 - Same instructions
 - Administered under same conditions
 - With appropriate accommodations for people with disabilities
 - Scored the same way for all test takers
- Not just multiple-choice questions:
 - Speaking tasks
 - Simulations
 - Performance-based activities

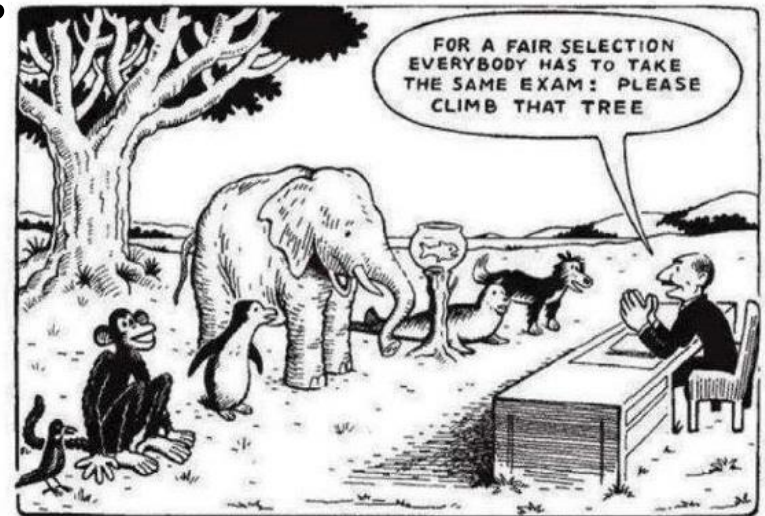
What is a “Standardized” Test?

Some examples of well-known standardized tests:

- OECD’s Program for International Student Assessment (PISA)
- The International Baccalaureate
- *TOEFL*® Test
- *GRE*® Tests

What is a “Standardized” Test?

- The fact that scales are standardized makes it possible to compare weights of different people or of the same person over time.
- Tests that are *not* standardized cannot be used to compare scores of different people or even the scores of the same person over time.



Value of Standardized Tests

- Fairness:
 - Designed, developed, and administered in ways that treat people equally and fairly regardless of differences in personal characteristics such as race, ethnicity, gender, or disability
- Reliability:
 - The extent to which scores (or other reported results) on a test are consistent across other forms of the test and, in some cases, other occasions of testing and other raters of the responses.
- Validity:
 - Programs should provide evidence to show that each test is capable of meeting its intended purposes
- Constructs Well Defined
 - Requires that the construct that will form the basis of an assessment must be clearly specified

Value of *TOEFL*® and *GRE*® Test Scores

- **TOEFL Test**

- **Measures all four language skills:** Reading, Listening, Writing and Speaking
- **Emphasizes the ability to use English** for communication
- **Uses integrated tasks** to measure how well students can use English in the classroom
- **Represents the tasks** typical of a higher education setting






- **GRE General Test**

- **Measures skills considered essential to graduate school success** by faculty and graduate deans
- **Provides the only common measures of these skills** for comparing the qualifications of applicants with different educational and cultural backgrounds
- **Furnishes independent information** to supplement the evaluation of grades and recommendations
- **Offers a valid predictor of graduate school performance** as confirmed by an independent study using 82,000 graduate students and over 1,700 studies containing validity data for *GRE*® tests

The background consists of several colored rectangular blocks. A blue block is in the top-left corner. A large yellow block covers the top-right and bottom-left areas. A dark blue block covers the middle section where the text is located. A maroon block is in the bottom-right corner.

The TOEFL® Testing Program:
Assessing Academic English Skills

The TOEFL iBT® Test Format

Section	Description	Testing Time	Score Scale
Reading 	3–4 passages from academic texts; 12–14 questions per passage	60–80 min.	0–30
Listening 	4–6 lectures, some with classroom discussion, 6 questions each; 2–3 conversations, 5 questions each	60–90 min.	0–30
Break 	-	10 min.	-
Speaking 	2 independent tasks – expressing an opinion; 4 integrated tasks – responding to questions based on reading and listening passages	20 min.	0–30
Writing 	1 integrated task – responding to reading and listening passages; 1 independent task – supporting an opinion on a topic	50 min.	0–30
Total	-	-	0–120

The Association between TOEFL iBT® Test Scores and the Common European Framework of Languages (CEFR) Levels

CEFR level	Reading (0–30)	Listening (0–30)	Speaking (0–30)	Writing (0–30)	Total (0–120)
C1 or above	24	22	25	24	95
B2	18	17	20	17	72
B1	4	9	16	13	42
A2	n/a	n/a	10	7	n/a

Research supporting the use of TOEFL in academic admissions

More than 240 peer-reviewed research reports, books, journal articles and book chapters to support test design, validity reliability and .

TOEFL iBT™ Research

Insight

Series I, Volume 4

Validity Evidence Supporting the Interpretation and Use of TOEFL iBT™ Scores

TOEFL iBT™ Research

Insight

Series I, Volume 3

Reliability and Comparability of TOEFL iBT™ Scores



Predicting grades from an English language assessment: The importance of peeling the onion

Language Testing
2016, Vol. 33(3) 307-318
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0265532215583066
ltj.sagepub.com
SAGE

Brent Bridgeman

Center for Foundational and Validity Research, Educational Testing Service, USA

Yeonsuk Cho

Center for English Language Learning and Assessment, Educational Testing Service, USA

Stephen DiPietro

Drexel University, USA

Abstract

Data from 787 international undergraduate students at an urban university in the United States were used to demonstrate the importance of separating a sample into meaningful subgroups in order to demonstrate the ability of an English language assessment to predict the first-year grade point average (GPA). For example, when all students were pooled in a single analysis, the correlation of scores from the Test of English as a Foreign Language (TOEFL) with GPA was .18; in a subsample of engineering students from China, the correlation with GPA was .58, or .77 when corrected for range restriction. Similarly, the corrected correlation of the TOEFL Reading score with GPA for Chinese business students changed dramatically (from .01 to .36) when students with an extreme discrepancy between their receptive (reading/listening) and productive (speaking/writing) scores were trimmed from the sample.

The background consists of several colored rectangular blocks. A blue block is in the top-left corner. A large yellow block covers the top-right and bottom-left areas. A dark blue block covers the middle section where the text is located. A maroon block is in the bottom-right corner.

The GRE[®] Testing Program:
Assessing Master Level Skills

The Computer-Delivered *GRE*® General Test

Section	Skills Measured	Number of Questions	Time
Analytical Writing (1 section)	Critical thinking and analytical writing skills, specifically the ability to articulate complex ideas clearly and effectively	1 "Analyze an Issue" task	30 mins.
		1 "Analyze an Argument" task	30 mins.
Verbal Reasoning (2 sections)	Reading comprehension, verbal and analytical reasoning skills, focusing on your ability to analyze and evaluate written material	20 questions per section	30 mins. per section
Quantitative Reasoning (2 sections)	Problem-solving ability, focusing on basic concepts of arithmetic, algebra, geometry and data analysis	20 questions per section	35 mins. per section
Unscored*	--	Varies	Varies
Research**	--	Varies	Varies

*An unidentified unscored section that does not count toward a score may be included and may appear in any order after the Analytical Writing section.

**An identified research section may be included in place of the unscored section. Questions in this section are included for ETS research purposes and will not count toward your score.

Validity

Validity research and analyses establish that the test measures what it is supposed to measure. The types of validity support that the *GRE*® Program has documented include:

- **Construct validity** (the test measures the skills/abilities that should be measured)
- **Content validity** (the test measures appropriate content)
- **Predictive validity** (the test predicts success)
- **Consequential validity** (the test demonstrates that adverse consequences are minimal)
- **External validity** (the test has the expected relationship with other measures of the same construct)

Although ETS works to accumulate validity evidence at each stage of the delivery and scoring process, the initial direction for validity research derives from feedback from members of the graduate and business school community, who provide information about the skills and abilities that they consider essential for success in graduate and business school.

Independent *GRE*® Validity Research

- Nathan R. Kuncel, Sarah A. Hazlett and Deniz S. Ones conducted a meta-analysis of all available GRE validity studies that had been conducted by independent researchers over several years
 - 1,753 independent samples of students
 - Over 80,000 graduate students
- Refinements in study included
 - A very large database of studies
 - Statistical corrections for restriction of range
 - Statistical corrections for unreliable criterion measures
- General Test was found to be a “generalizably valid predictor” of
 - First-year graduate grade point average
 - Overall graduate grade point average
 - Comprehensive exam scores
 - Publication citation counts
 - Faculty ratings
- Correlations with degree attainment and research productivity were consistently positive, but not statistically significant in all cases

Results of Kuncel et al. Research

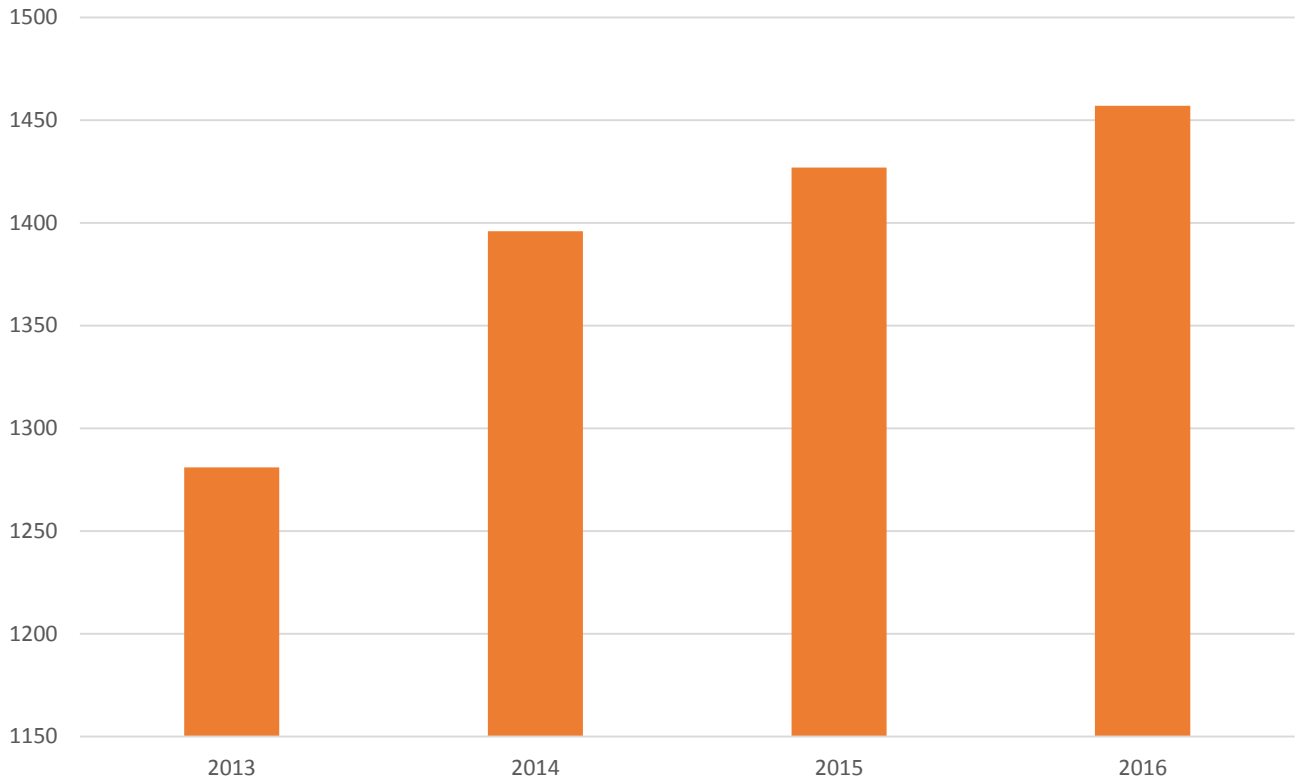
- A follow-up study by Kuncel indicated the *GRE*[®] General Test has similar levels of predictive validity for both master's- and doctoral-level programs
- The GRE General Test has better predictive validity than undergraduate grades or letters of recommendation
- The Subject Tests were found to be better predictors of success than were either the General Test or undergraduate grade point average

These data strongly suggest that the GRE tests are valid predictors of graduate student performance across disciplines, degree levels, age of students, etc.

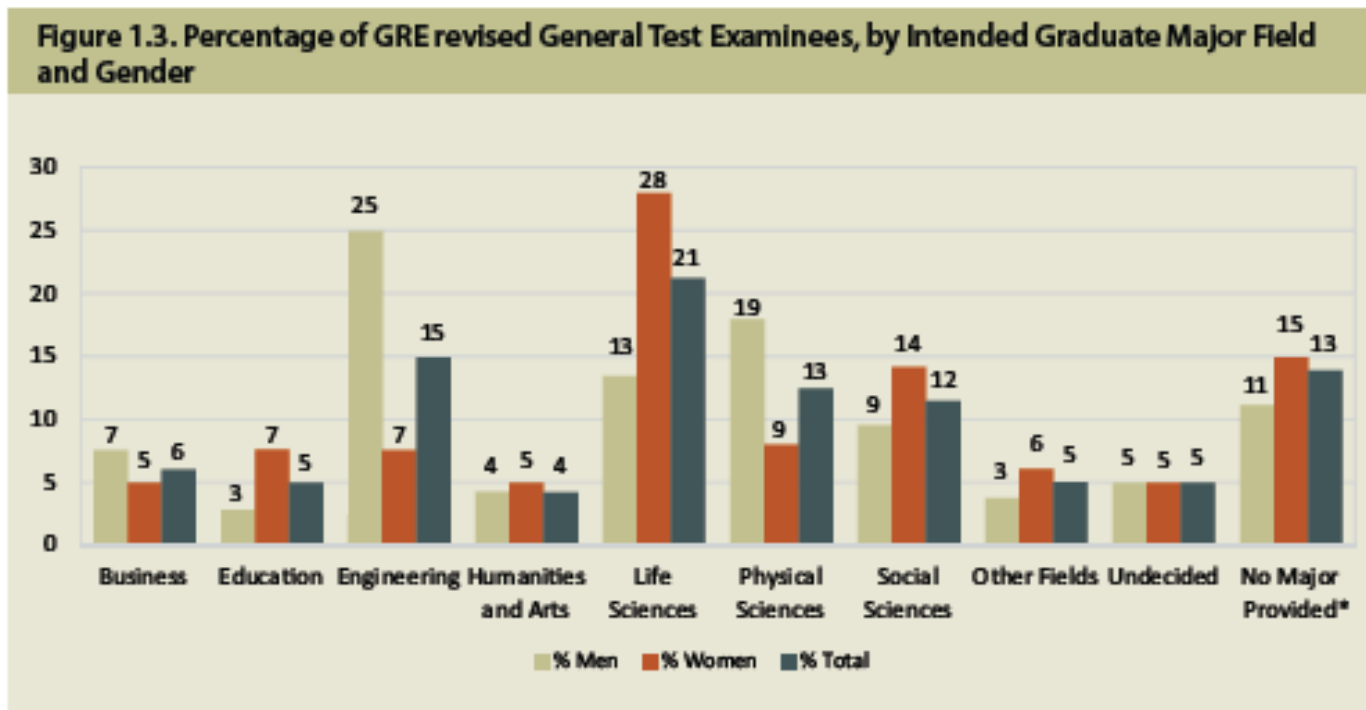
For more information about the Kuncel, et al. research study, see *A Comprehensive Review of Published GRE[®] Validity Data* available at www.ets.org/gre/institutions/about/downloads/

The use of the *TOEFL*® and *GRE*®
Testing Programs in Europe:
Growing Numbers

Growing Number of European Institutions Using TOEFL Scores



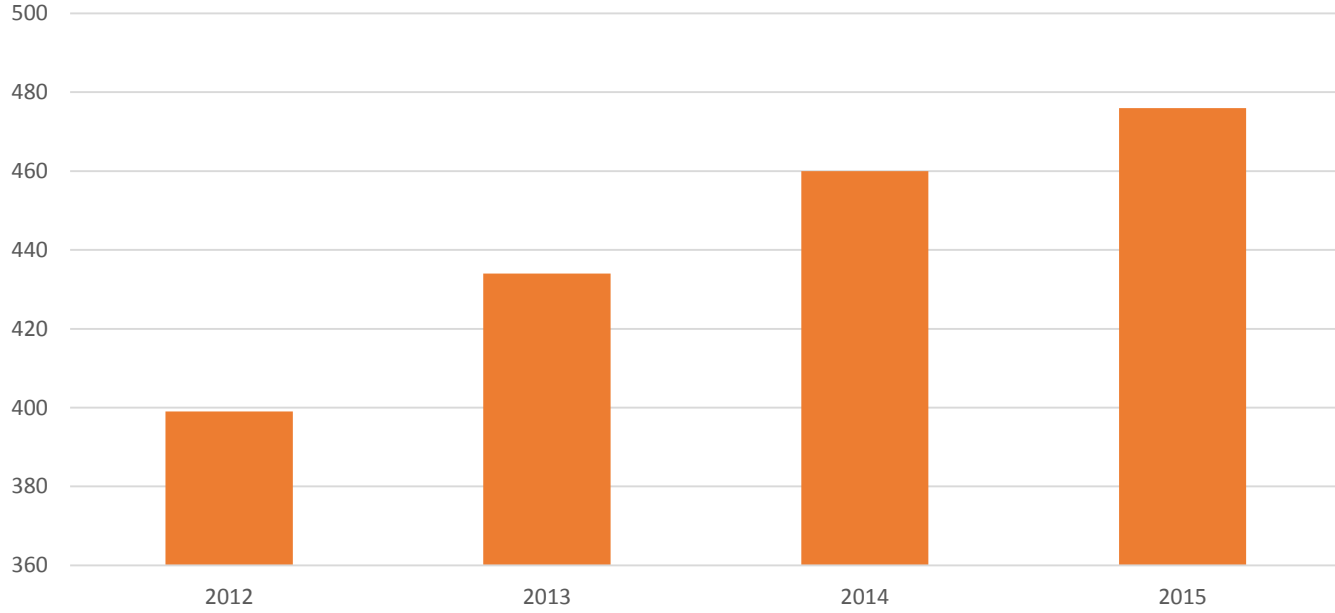
Intended Field of Study of the GRE® Test Takers (from the “Snapshot”)



Note: See page 23 for a list of major fields. These analyses were based on a total of 576,220 test takers. The percentages in the figure above represent the percentage of test takers within a particular gender category (e.g., men, women) for each intended graduate major field.

*Test takers in the “No Major Provided” category include those who indicated their intended graduate major as “Any Department Not Listed,” those who did not respond to the question, and those who provided an invalid answer.

Growing Number of European Institutions Using GRE scores



The Use of the *TOEFL*® and *GRE*®
Testing Programs in Europe:
Best Practices & Data

Admissions – Best Practices

- Check your score requirements regularly and adjust as needed
 - Each university sets its own score requirements
 - No 'passing' score for *TOEFL*® and *GRE*® tests
 - Benchmarking with similar institutions
- Evaluate an applicant's academic readiness on all available relevant information, not solely on test results
 - Grades/transcripts
 - Interviews,
 - Extracurricular activities
 - Professional experience,
 - etc.
- Use score ranges rather than rigid cut scores
 - Section scores (sub scores)
 - Standard Errors of Measurement

The *GRE*[®] Guidelines Support Holistic File Review

- The *Guide to the Use of GRE*[®] *Scores* recommends the following:
 - Multiple sources of information should be used to ensure fairness and to balance the limitations of any single measure of knowledge, skills and abilities
 - Use of multiple criteria is particularly important when assessing the abilities of educationally disadvantaged applicants, applicants whose primary language is not English and applicants who are returning to school after an extended absence
 - A cut score should never be used as the only criterion for denial of admission or awarding of a fellowship
 - Small differences in GRE scores (as defined by the standard error of measurement) should not be used to make distinctions among test takers

Using *TOEFL*® and *GRE*® Data in the Admission and Recruitment Processes

Data Collection and Research

- Annual CGS/GRE Survey of Graduate Enrollment
- Access to Research Reports
- GRE® Snapshot & TOEFL annual data report

Admissions

- ETS Data Manager
- Score Interpretation Information and Advice

Recruitment

- Search Services for TOEFL® and GRE® (500,000+ test takers)
- Promotion of Recruitment Events on Facebook and Weibo

Questions?

zkaptan@etsglobal.org