Foreword

We are very excited to announce the TOEFL iBT® Research Insight Series, a bimonthly publication to make important research on the TOEFL iBT available to all test score users in a user-friendly format.

The TOEFL iBT test is the most widely accepted English language assessment, used for admissions purposes in more than 130 countries including the United Kingdom, Canada, Australia, New Zealand and the United States. Since its initial launch in 1964, the TOEFL® test has undergone several major revisions motivated by advances in theories of language ability and changes in English teaching practices. The most recent revision, the TOEFL iBT test, was launched in 2005. It contains a number of innovative design features, including the use of integrated tasks that engage multiple language skills to simulate language use in academic settings, and the use of test materials that reflect the reading and listening demands of real-world academic environments.

At ETS we understand that you use TOEFL iBT test scores to help make important decisions about your students, and we would like to keep you up to date about the research results that assure the quality of these scores. Through the TOEFL iBT Research Insight Series we wish to both communicate to the institutions and English teachers who use the TOEFL iBT test scores the strong research and development base that underlies the TOEFL iBT test, and demonstrate our strong, continued commitment to research.

We hope you will find this series relevant, informative and useful. We welcome your comments and suggestions about how to make it a better resource for you.

Ida Lawrence
Senior Vice President,
Research & Development Division,
Educational Testing Service

Preface

Since the 1970’s, the TOEFL test has had a rigorous, productive and far-ranging research program. But why should test score users care about the research base for a test? In short, because it is only through a rigorous program of research that a testing company can demonstrate its forward-looking vision and substantiate claims about what test takers know or can do based on their test scores. This is why ETS has made the establishment of a strong research base a consistent feature of the evolution of the TOEFL test.

The TOEFL test is developed and supported by a world-class team of test developers, educational measurement specialists, statisticians and researchers. Our test developers have advanced degrees in such fields as English, language education and linguistics. They also possess extensive international experience, having taught English in Africa, Asia, Europe, North America and South America. Our research, measurement and statistics team includes some of the world’s most distinguished scientists and internationally recognized leaders in diverse areas such as test validity, language learning and testing, and educational measurement and statistics.

To date, more than 150 peer-reviewed TOEFL research reports, technical reports and monographs have been published by ETS, many of which have also appeared in academic journals and book volumes. In addition to the 20-30 TOEFL-related research projects conducted by ETS Research & Development staff each year, the TOEFL Committee of Examiners (COE), comprised of language learning and testing experts from the academic community, funds an annual program of TOEFL research by external researchers from all over the world, including preeminent researchers from Australia, the UK, the US, Canada and Japan.

In Series One of the TOEFL iBT Research Insight Series, we provide a comprehensive account of the essential concepts, procedures and research results that assure the quality of scores on the TOEFL iBT test. The six issues in this Series will cover the following topics:
Issue 1: TOEFL iBT® Test Framework and Development

The TOEFL iBT® test is described along with the processes used to develop test questions and forms. These processes include rigorous review of test materials, with special attention to fairness concerns. Item pretesting, try outs and scoring procedures are also detailed.

Issue 2: TOEFL Research

The TOEFL® Program has supported rigorous research to maintain and improve test quality. Over 150 reports and monographs are catalogued on the TOEFL website. A brief overview of some recent research on fairness and automated scoring is presented here.

Issue 3: Reliability and Comparability of Test Scores

Given that hundreds of thousands of test takers take the TOEFL iBT® test each year, many different test forms are developed and administered. Procedures to achieve score comparability on different forms are described in this section.

Issue 4: Validity Evidence Supporting Test Score Interpretation and Use

The many types of evidence supporting the proposed interpretation and use of test scores as a measure of English-language proficiency in academic contexts are discussed.

Issue 5: Information for Score Users, Teachers and Learners

Materials and guidelines are available to aid in the interpretation and appropriate use of test scores, as well as resources for teachers and learners that support English-language instruction and test preparation.

Issue 6: TOEFL Program History

A brief overview of the history and governance of the TOEFL Program is presented. The evolution of the TOEFL test constructs and contents from 1964 to the present is summarized. Future series will feature summaries of recent studies on topics of interest to our score users, such as “what TOEFL iBT test scores tell us about how examinees perform in academic settings,” and “how score users perceive and use TOEFL iBT test scores.”

The close collaboration with TOEFL iBT score users, English language learning and teaching experts and university professors in the redesign of the TOEFL iBT test has contributed to its great success. Therefore, through this publication, we hope to foster an even stronger connection with our score users by sharing the rigorous measurement and research base and solid test development that continues to ensure the quality of TOEFL iBT scores meets the needs of score users.

Contributors

The primary author of this section is Mary Enright.

The following individuals also contributed to this section by providing their careful review as well as editorial suggestions (in alphabetical order):

- Terry Axe
- Cris Breining
- Brent Bridgeman
- Don Powers
- Rosalie Szabo
- Mikyung Kim Wolf
- Xiaoming Xi

History of the TOEFL® Program

Origins and Governance of the TOEFL Program

The TOEFL® test, formerly known as the “Test of English as a Foreign Language” was developed in the early 1960’s to assess the English proficiency of nonnative speakers of the language who intend to study in institutions where English is the language of instruction. The TOEFL test was developed under the auspices of the National Council on the Teaching of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private institutions concerned with the English proficiency of nonnative speakers, especially those applying to English-medium academic institutions. The testing program was initially financed by grants from the Ford and Danforth Foundations and was, at first, attached administratively to the Modern Language Association. In 1965, the College Board and Educational Testing Service (ETS) assumed joint responsibility for the testing program.

In recognition of the fact that many who take the TOEFL test are potential graduate students, a cooperative arrangement for the operation of the testing program was entered into by ETS, the College Board, and the Graduate Record Examinations® (GRE®) Board.

Today the TOEFL Board advises ETS on policies under which ETS administers the TOEFL Program. The Board consists of 16 individuals. In accordance with its bylaws, the College Board appoints three Board members, and three are appointed by the GRE Board to represent the interests of undergraduate and graduate education. Nine at-large members are also appointed to provide representation for the various constituent groups that have an interest in the program. The Chair of the TOEFL Committee of Examiners also serves on the TOEFL Board.

The board has three standing committees, each responsible for specific areas of program activity: the Committee of Examiners (COE), the Grants and Awards Committee and the Finance Committee.

The Committee of Examiners is a standing committee of the TOEFL Board. The COE consists of twelve individuals from around the world, each of whom has achieved professional recognition in an academic field related to English as a Foreign or Second Language. Members provide guidance and oversight for research and development related to the TOEFL test. On other matters, such as major test redesign and policy changes, the COE has power of recommendation to the TOEFL Board. In addition, the COE has a Research Subcommittee responsible for commissioning and overseeing research carried out under the COE research program.

The Board has a Grants and Awards Committee that oversees eight categories of awards and grants annually in the fields of language teaching, testing and research and international education.

Finally, the Board also has a standing Finance Committee.

The Evolution of the TOEFL Test

Since its inception in 1963, the TOEFL test has evolved from a paper-based test, to a computer-based test and, in 2005, to an internet-based test. But more important than changes in the method of test delivery are the changes in the theories of language proficiency that motivated the redesign of the test’s construct and content.

Since its inception in 2002, the TOEFL test has evolved from a paper-based test, to a computer-based test and, in 2005, to an internet-based test. But more important than changes in the method of test delivery are the changes in the theories of language proficiency that motivated the redesign of the test’s construct and content. In Table 1 we describe the construct underlying the test at three stages of its evolution and how it influenced test content.

Taylor and Angelis (2008) provide a detailed description of how thinking in the fields of linguistics and educational psychology and measurement influenced the development of the TOEFL test at each of these stages.
Table 1. The Evolution of the TOEFL Test Constructs and Contents over Three Stages of Development.

<table>
<thead>
<tr>
<th>Stages</th>
<th>Construct</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The First TOEFL Test 1964–1979</td>
<td>Discrete components of language skills and knowledge</td>
<td>Multiple-choice items assessing vocabulary, reading comprehension, listening comprehension, knowledge of correct English structure and grammar</td>
</tr>
<tr>
<td>2. A Suite of TOEFL Tests 1979–2005</td>
<td>Original constructs (listening, reading, structure and grammar) retained but two additional ones added—writing ability and speaking ability</td>
<td>In addition to multiple-choice items assessing the original constructs, separate constructed-response tests of writing, the TWE test and speaking, the TSE test, were developed</td>
</tr>
<tr>
<td>3. The TOEFL iBT® Test 2005–present</td>
<td>Communicative competence—the ability to put language knowledge to use in relevant contexts</td>
<td>Academic tasks were developed that require the integration of receptive and productive skills such as listening, reading and writing, or speaking, as well multiple-choice items for listening and reading</td>
</tr>
</tbody>
</table>

1. The First TOEFL Test. When the first TOEFL test was developed, linguistic theories envisioned language competence as a set of separate components such as grammar, vocabulary and comprehension (Carroll, 1961; Lado, 1961). This focus on discrete language components or skills was compatible with multiple-choice measurement practices that were current at the time. At this early stage, the need to assess other important skills was recognized, although methods to do so on a large scale were not yet available. These other important skills included speaking, writing and “integrative” skills (Carroll, 1961) that required test takers to combine discrete language components in the service of communication—the construction and production of meaning. While the first TOEFL test did not assess these skills, research and development continued over the following years with the goal of incorporating these skills in future assessments.

2. A Suite of TOEFL Tests. In the 1970’s, the second major stage of TOEFL development, attention turned to the expansion of the TOEFL construct to include speaking and writing. A multiple-choice TOEFL test that assessed reading, listening, and structure and written expression continued to be administered to all candidates. ETS also developed separate tests of speaking and writing. Initially these tests were taken by smaller numbers of applicants to fulfill the requirements of specific institutions. The TSE® Test, formerly known as the Test of Spoken English™, was developed primarily to assess the oral skills of international graduate students who were being considered for positions as teaching assistants. On the TSE test, test takers tape-recorded oral responses to prerecorded questions that targeted specific language functions. The TWE® Test, formerly known as the Test of Written English™, introduced in 1986, consisted of a single essay written in response to a brief question or topic. The TWE test required test takers to develop and organize ideas and express these ideas using appropriate English vocabulary and grammar. Initially, the TWE test was administered as a required component of the TOEFL test at five administrations per year.

During this second stage of the TOEFL test development, the technology to administer tests was also changing. And language experts continued to elaborate on Carroll’s ideas about “integrative” language skills, and theories of “communicative competence” emerged (Bachman & Palmer, 1996; Canale & Swain, 1980; Hymes, 1972). Throughout this period, ETS sponsored studies to explore the relationship between theories of “communicative competence” and the TOEFL test (Duran, Canale, Penfield, Stansfield, & Liskin-Gasparro, 1985; Henning & Cascallar, 1992).

3. The TOEFL iBT Test. In the 1990’s the TOEFL program, in consultation with the COE, considered how to design a new TOEFL test that was more reflective of communicative competence. The construct of communicative competence that would guide the design of the new test was elaborated further as communicative language use in academic contexts (Chapelle, Grabe, & Berns, 1997). In academic contexts, language use engages both receptive (listening, reading) and productive (speaking, writing) modalities. And many academic tasks require the use of these modalities in a complementary or integrated manner.

Beginning in 1996, teams were formed and charged with the responsibility of designing a new TOEFL test that embodied the construct of communicative competence. These teams were composed of ETS staff and external advisers with expertise in test development, applied linguistics, research methods and psychometrics. The test design process began with the development of a general framework that stated that the new “...test will measure examinees’ English-language proficiency in situations and tasks reflective of university life...” (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000, p.11). Subsequently, working papers were written that extended this conceptual framework to four language domains—listening (Bejar, Douglas, Jamieson, Nissen, & Turner, 2000), reading (Ennright, Grabe, Koda, Mosenthal, Mulcahy-Errt, & Schell, 2000), speaking (Butler, Eignor, Jones, McNamara, & Suomi, 2000) and writing (Cumming, Kantor, Powers, Santos & Taylor, 2000). Each of these working papers defined processes for designing tasks that could assess test takers’ proficiency within the relevant domain and included a research agenda. Each research agenda laid out plans to develop and evaluate new tasks and measures and to gather the kinds of evidence that would support the valid interpretation and use of the new test scores. The extent of the program of research carried out to support the design of the TOEFL iBT test is documented in a series of series of 35 monographs (ETS, 2009) and a book (Chapelle, Enright, & Jamieson, 2008).

Evidence-centered design (ECD), a test design methodology developed at ETS, was used to finalize the test specifications for each of the four main skills assessed in the TOEFL iBT test and all of the relevant subskills (Pearlman, 2008). An important part of the development of the test specifications was to ensure that the associated test questions were comparable across tasks and that they could be developed in sufficient numbers to sustain ongoing test development.

For more than 40 years the TOEFL program has pioneered innovations in language assessment that were consistent with advances in our understanding of language competence and with developments in technology.

ETS sponsored studies to explore the relationship between theories of “communicative competence” and the TOEFL test.
TOEFL Program History

References


