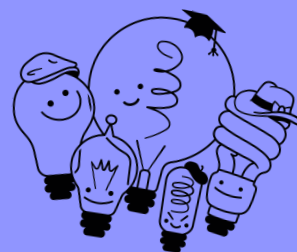




# **TOEFL® Research INSIGHTS**

## **TOEFL® Program History**

**VOLUME 6**



# TOEFL® Research Insight Series, Volume 6: TOEFL Program History

## Preface

The TOEFL iBT® test is the world's most widely respected English language assessment and used for admissions purposes in more than 150 countries, including Australia, Canada, New Zealand, the United Kingdom, and the United States (see test review in Alderson, 2009). 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 integrated tasks that engage multiple skills to simulate language use in academic settings and test materials that reflect the reading, listening, speaking, and writing demands of real-world academic environments.

In addition to the TOEFL iBT test, the TOEFL® Family of Assessments was expanded to provide high-quality, English proficiency assessments for a variety of academic uses and contexts. The TOEFL® Young Students Series features the TOEFL Primary® and TOEFL Junior® tests, which are designed to help teachers and learners of English in school settings. In addition, the TOEFL ITP® program offers colleges, universities, and others affordable tests for placement and progress monitoring within English programs as a pathway to eventual degree programs.

At ETS, we understand that scores from the TOEFL Family of Assessments are used to help make important decisions about students, and we would like to keep score users and test takers up-to-date about the research results that help assure the quality of these scores. Through the publication of the TOEFL® Research Insight Series, we wish to communicate to institutions and English teachers who use the TOEFL tests the strong research and development base that underlies the TOEFL Family of Assessments and demonstrate our continued commitment to research.

Since the 1970s, 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, it is only through a rigorous program of research that a testing company can substantiate claims about what test takers know or can do based on their test scores, as well as provide support for the intended uses of assessments and minimize potential negative consequences of score use. Beyond demonstrating this critical evidence of test quality, research is also important for enabling innovations in test design and addressing the needs of test takers and test score users. This is why ETS has established a strong research base as a fundamental feature underlying the evolution of the TOEFL Family of Assessments.

This portfolio is designed, produced, and supported by a world-class team of test developers, educational measurement specialists, statisticians, and researchers in applied linguistics and language testing. Our test developers have advanced degrees in fields such as English, language education, and applied linguistics. They also possess extensive international experience, having taught English on continents around the globe. Our research, measurement, and statistics teams include some of the world's most distinguished scientists and internationally recognized leaders in diverse areas such as test validity, language learning and assessment, and educational measurement.

To date, more than 300 peer-reviewed TOEFL Family of Assessments research reports, technical reports, and monographs have been published by ETS, and many more studies on the TOEFL tests have appeared in academic journals and book volumes. In addition, over 20 TOEFL test-related research projects are conducted by ETS's Research & Development staff each year and the TOEFL Committee of Examiners — comprising language learning and testing experts from the global academic community — funds an annual program of TOEFL Family of Assessments research by independent external researchers from all over the world.

The purpose of the *TOEFL Research Insight Series* is to provide a comprehensive, yet user-friendly account of the essential concepts, procedures, and research results that assure the quality of scores for all products in the TOEFL Family of Assessments. Topics covered in these volumes feature issues of core interest to test users, including how tests were designed; evidence for the reliability, validity, and fairness of test scores; and research-based recommendations for best practices.

The close collaboration with TOEFL test score users, English language learning and teaching experts, and university scholars in the design of all TOEFL tests has been a cornerstone to their success and worldwide acceptance. Therefore, through this publication, we hope to foster an ever-stronger connection with our test users by sharing the rigorous measurement and research base, as well as solid test development, that continues to help ensure the quality of the TOEFL Family of Assessments.

John Norris, Ph.D.

Senior Research Director

English Language Learning and Assessment

Research & Development Division

ETS



## Acknowledgements

The following ETS staff contributed to this version of Volume 6, updated in November 2024 (in alphabetical order): Larry Davis, Michelle Hampton, Spiros Papageorgiou.

The following individuals also contributed to previous versions of this volume, by providing careful reviews and revisions, as well as editorial suggestions (in alphabetical order): Terry Axe, John Norris, Eileen Tyson, Jennifer Wain, and Yuan Wang. The primary author of the first edition was Mary Enright. Terry Axe, Cristiane Breining, Brent Bridgeman, Don Powers, Rosalie Szabo, Mikyung Kim Wolf, and Xiaoming Xi also contributed to the first edition.

# History of the TOEFL Program

## Origins and Governance of the TOEFL Program

The TOEFL test was developed in the early 1960s to assess the English proficiency of second language speakers of English who intend to study in institutions where English is the language of instruction. The TOEFL test was commissioned under the auspices of the National Council on the Testing of English as a Foreign Language, a temporary working group formed through the cooperative effort of more than thirty public and private institutions concerned with the English proficiency of second language speakers, especially those applying to English-medium academic institutions. The TOEFL testing program was initially financed by grants from the Ford and Danforth Foundations and was attached administratively to the Modern Language Association. In 1965, the College Board® and ETS assumed joint responsibility for the testing program.

In recognition of the fact that many who take the TOEFL test are potential undergraduate and 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. In 1973, the TOEFL Board was created to advise ETS on policies under which the TOEFL program is administered. The TOEFL Board consists of individuals that represent the interests of undergraduate education, graduate education and various other constituent groups that have an interest in the program.

The TOEFL Committee of Examiners (COE) is a standing committee of the TOEFL Board. The COE consists of individuals from around the world, all experts in the fields of English language teaching, testing, and research, who provide guidance and oversight for research and development related to the TOEFL iBT test as well as other examinations in the TOEFL Family of Assessments. The chair of the COE also serves on the TOEFL Board. On other matters, such as major test redesign and policy changes, the COE has power of recommendation to the TOEFL Program leadership. In addition, the COE has three research subcommittees. One subcommittee is responsible for commissioning and overseeing research focusing on the adult-oriented TOEFL iBT and TOEFL ITP assessments. A second subcommittee is responsible for commissioning and overseeing research on the TOEFL Young Students assessments (TOEFL Primary and TOEFL Junior tests). A third subcommittee is responsible for the English-language Researcher/Practitioner Grant Program (<https://www.ets.org/toefl/grants.html>), including selection of winners each year and support of the projects they conduct.

# The Evolution of the TOEFL Test

Since its inception in 1964, the TOEFL test has evolved from a paper-based test to a computer-based test and, in 2005, to an internet-based test. More importantly, beyond changes in the medium of test delivery, innovations in theories of language proficiency have motivated the redesign of the test’s construct (what it measures), content, and item types. Table 1 describes the evolving construct underlying the test during three stages of its development, as well as how these changes were reflected in test content and item types.

For further information, Taylor and Angelis (2008) provided a detailed description of how thinking in the fields of applied linguistics and educational psychology and measurement influenced the development of the TOEFL test at each of these stages.

Since its inception in 1964 the *TOEFL test* has evolved from a paper-based test to a computer-based test and, in 2005, to an internet-based test. More importantly, *beyond changes in the medium of test delivery, innovations in theories of language proficiency* have motivated the redesign of the test’s construct (what it measures), content, and item types.

**Table 1. The Evolution of the TOEFL Test Construct and Content Over Three Stages of Development**

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

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 item types, paper-and-pencil testing, and 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 order to communicate — the construction and production of meaning. While the first TOEFL test did not assess these integrative skills, research and development continued over the following years with the goal of incorporating them in future assessments.
2. **A Suite of TOEFL Tests.** In the 1970s, 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. Additionally, in recognition of the need to assess an actual ability to communicate in English, ETS 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, originally 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, examinees tape-recorded oral responses to prerecorded questions that targeted specific language functions, and their responses were later scored by trained raters using the TSE rating rubric. The TWE test, originally known as the Test of Written English, was introduced in 1986 and 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 but only offered at five select administrations per year.

During this second stage of TOEFL test development, the technology used to administer tests was also changing. Toward the end of this stage, in 1998, a computer-based version of the TOEFL test (TOEFL CBT), which included a mandatory essay component for all test takers, was introduced. Furthermore, language experts continued to elaborate on Carroll's (1961) ideas about integrative language skills, and theories of communicative competence emerged that stressed the fundamentally communicative nature of language use (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). The development of the TOEFL CBT test and the continued interest in communicative competence signaled a transition to the next stage in the evolution of the TOEFL test.

*...ETS sponsored studies to explore the relationship between theories of communicative competence and the TOEFL test.*

3. The TOEFL iBT Test. In the 1990s, the TOEFL program, in consultation with the COE, considered how to design a new TOEFL test that was more reflective of communicative competence and that could be delivered efficiently to an expanding variety of test takers worldwide while maintaining key aspects of standardized test administration, scoring, and security. The optimal delivery medium for the test was determined to be via the internet at official, authorized test centers, as the time between taking a test and receiving a score could be reduced considerably. Hence, the original name of the test was the TOEFL internet-based test, later abbreviated to TOEFL iBT.

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) abilities, and many academic tasks require the use of these abilities in a complementary or integrated manner.

Beginning in 1996, teams of experts were formed and charged with the responsibility of designing a new TOEFL test that embodied this construct of communicative language use in academic contexts. 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 founded on the notion that the new test would measure “examinees’ English language proficiency in situations and tasks reflective of university life” (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000, p. 11).

Two important frameworks have guided test design and validation efforts for the TOEFL iBT test. 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 contribution of these 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. ECD provides the essential basis for delivering a standardized test that is reliable and fair across numerous test forms and administrations.

Subsequently, working papers were written that extended this conceptual framework to four language domains — listening (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000), reading (Enright et al., 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 test items that would reflect academic tasks and assess test takers' proficiency within the relevant language domain. They also established a research agenda that laid out plans to develop and evaluate new tasks and to gather the kinds of evidence that would support the valid interpretation and use of the new test scores. The associated program of research carried out to support the full design of the TOEFL iBT test is documented in a series of 35 monographs and a book (Chapelle, Enright, & Jamieson, 2008).

Validity research for the TOEFL iBT test has been guided by an argument-based approach (Kane, 2013) that helps to lay out the different assumptions or claims explaining how the test is supposed to work to provide meaningful information about a test taker's academic communicative competence in English. It also establishes the types of evidence needed to support these claims. Initial validity evidence for the TOEFL iBT test is compiled according to the argument-based approach in the book edited by Chapelle, Jamieson, and Enright (2008).



## The Expanding TOEFL Family of Assessments

While most people worldwide associate the TOEFL name with the TOEFL iBT test specifically, the TOEFL brand name actually represents a family of assessments designed to meet the needs of different test users.

In addition to the TOEFL iBT test, several versions of the TOEFL test are offered for distinct purposes:

- The TOEFL® Essentials™ test measures foundational language skills and communication abilities in academic and general (daily life) contexts. The test covers the four language skills of reading, listening, writing, and speaking in a shorter and accessible format and is intended to provide academic programs and other users with reliable information regarding the test taker's ability to understand and use English.
- TOEFL ITP Assessment Series program: Many educational programs at universities and other institutions desire an assessment of students' English language proficiency for making placement decisions, for monitoring progress, and for exiting students from second language instruction. For these purposes, the TOEFL ITP test was created as a flexible, low-cost, and relatively low-stakes test that does not require the same kinds of security measures, fixed administration schedule, or administration time demands as the TOEFL iBT test. The TOEFL ITP test is administered locally and scores are reported to and used within the administering institution.

Beyond these assessments that are focused on adult learners and university-level academic communication, several additional assessments have been developed more recently for use with younger learners of English:

- In 2010, ETS launched the TOEFL Junior tests as age-appropriate tests of academic English for learners ages 11 and up. The TOEFL Junior tests are available in a variety of formats and evaluate students' abilities to use English for communication in middle and high school environments. The tests are used by schools and teachers for placement in language programs, monitoring learner progress, and making related decisions.
- In 2013, the TOEFL Primary tests were launched as a classroom-oriented assessments for English learners ages 8 and up. The main purpose of the TOEFL Primary tests is the provision of feedback to teachers and learners as a basis for guiding their educational efforts.

## Summary

For more than 40 years, the TOEFL program has pioneered innovations in language assessment that are consistent with advances in understanding about language competence and with developments in technology. The TOEFL program has also sought to respond to the needs of test users by continuing to develop alternative administration formats for its assessments and by developing new assessments for English learners at different ages. Accompanying these innovations has been a persistent commitment to the highest standards of educational measurement practice (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), continuous audits of test programs following the ETS Standards for Quality and Fairness (Educational Testing Service, 2014), and, in particular, a sustained program of validation research.

*For more than 40 years, the TOEFL program has pioneered innovations in a language assessment that are consistent with advances in understanding about language competence and with developments in technology.*

# References

- Alderson, J. C. Test review: Test of English as a Foreign Language™: Internet-based Test (TOEFL iBT®). *Language Testing*, 26(4), 621-631. doi:10.1177/0265532209346371
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper* (TOEFL Monograph No. 19). Princeton, NJ: Educational Testing Service.
- Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph No. 20). Princeton, NJ: Educational Testing Service.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- Carroll, J. B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In *Testing the English proficiency of foreign students* (pp. 31–40). Washington, DC: Center for Applied Linguistics.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.
- Chapelle, C., Grabe, W., & Berns, M. (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000* (TOEFL Monograph No. 10). Princeton, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., Powers, D. E., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph No. 18). Princeton, NJ: Educational Testing Service.
- Duran, R., Canale, M., Penfield, J., Stansfield, C., & Liskin-Gasparro, J. (1985). *TOEFL from a communicative viewpoint on language proficiency: A working paper* (TOEFL Research Report No. 17). Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (2014). *ETS standards for quality and fairness*.

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph No. 17). Princeton, NJ: Educational Testing Service.

Henning, G., & Cascallar, E. (1992). *A preliminary study of the nature of communicative competence* (TOEFL Research Report No. 36). Princeton, NJ: Educational Testing Service.

Hymes, D. (1972). On communicative competence. In B. J. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, England: Penguin.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper* (TOEFL Monograph No. 16). Princeton, NJ: Educational Testing Service.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.

Lado, R. (1961). *Language testing*. New York, NY: McGraw-Hill.

Pearlman, M. (2008). Finalizing the test blueprint. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 227–258). New York, NY: Routledge.

Taylor, C., & Angelis, P. (2008). The evolution of TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 27–54). New York, NY: Routledge.