

TOEFL® Research – Ensuring Test Quality

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, the TOEFL iBT™ test. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. In 1997, a monograph series that laid the groundwork for the development of the TOEFL iBT test was launched and research on this new test continues today. Currently this research is carried out in consultation with the TOEFL Committee of Examiners. Its members include representatives of the TOEFL Board and distinguished English-as-a-Second-Language specialists from the academic community.

Recent research reports that are currently available or are soon to be available are categorized below with respect to the many different types of evidence that speak to a test's quality. Links to reports are provided as they become available. You can click the title while holding down the control key to read the entire report.

Validity Evidence

Syntheses of Validity Evidence

Chappelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2008). [*Building a validity argument for the Test of English as a Foreign Language™*](#). New York: Routledge.

Educational Testing Service. (2011). [*Validity evidence supporting the interpretation and use of TOEFL iBT™ scores*](#) (TOEFL Research Insight Series Vol. 4). Princeton, NJ: ETS.

Relationship Among Test Measures

Sawaki, Y., Stricker, L. J., & Oranje, A. (2008). [*Factor structure of the TOEFL® Internet-based test \(iBT\): Exploration in a field trial sample*](#) (TOEFL iBT™ Report No. iBT-04). Princeton, NJ: ETS.

Stricker, L. J., Rock, D. A., & Lee, Y.-W. (2005). [*Factor structure of the LanguEdge™ test across language groups*](#) (TOEFL® Monograph No. MS-32). Princeton, NJ: ETS.

Stricker, L. J., & Rock, D. A. (2008). [*Factor structure of the TOEFL® Internet-based test across subgroups*](#) (TOEFL iBT™ Report No. iBT-07). Princeton, NJ: ETS.

Construct Representation

Biber, D. & Gray, B. (2013). [*Discourse characteristics of writing and speaking task types on the TOEFL iBT Test: A lexico-grammatical analysis*](#) (TOEFL iBT™ Report No. iBT-19). Princeton, NJ: ETS.

Brown, A., Iwashita, N., & McNamara, T. (2005). [*An examination of rater orientations and test taker performance on English-for-academic-purposes speaking tasks*](#) (TOEFL® Monograph No. MS-29). Princeton, NJ: ETS.

Cohen, A. D., & Upton, T. A. (2006). [*Strategies in responding to new TOEFL® reading tasks*](#) (TOEFL® Monograph No. MS-33). Princeton, NJ: ETS.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2005). [*Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL® test*](#) (TOEFL® Monograph No. MS-30). Princeton, NJ: ETS.

Swain, M., Huang, L., Barkaoui, K., Brooks, L., & Lapkin, S. (2009). [*The speaking section of the TOEFL iBT™ \(SSTiBT\): Test-takers' reported strategic behaviors*](#) (TOEFL iBT™ Report No. iBT-10). Princeton, NJ: ETS.

Criterion-Related & Predictive Validity

Powers, D. E., Roever, C., Huff, K. L., & Trapani, C. S. (2003). [*Validating LanguEdge™ courseware against faculty ratings and student self-assessments*](#) (ETS Research Rep. No. RR-03-11). Princeton, NJ: ETS.

Roever, C., & Powers, D. E. (2005). [*Effects of language of administration on a self-assessment of language skills*](#) (TOEFL® Monograph No. MS-27). Princeton, NJ: ETS.

Sawaki, Y., & Nissan, S. (2009). [*Criterion-related validity of the TOEFL® iBT listening section*](#) (TOEFL iBT™ Report No. iBT-08). Princeton, NJ: ETS.

Weigle, S. C. (2011). [*Validation of automated scores of TOEFL iBT® tasks against nontest indicators of writing ability*](#) (TOEFL iBT™ Report No. iBT-15). Princeton, NJ: ETS.

Xi, X. (2008). [*Investigating the criterion-related validity of the TOEFL® speaking scores for ITA screening and setting standards for ITAs*](#) (TOEFL iBT™ Report No. iBT-03). Princeton, NJ: ETS.

Authenticity and Content Validity

Biber, D., Conrad, S. M., Reppen, R., Byrd, P., Helt, M., Clark, V., et al. (2004). [*Representing language use in the university: Analysis of the TOEFL® 2000 spoken and written academic language corpus*](#) (TOEFL® Monograph No. MS-25). Princeton, NJ: ETS.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2005). [*A teacher-verification study of speaking and writing prototype tasks for a new TOEFL® test*](#) (TOEFL Monograph No. MS-26). Princeton, NJ: ETS.

Rosenfeld, M., Leung, S., & Oltman, P. K. (2001). [*The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels*](#) (TOEFL® Monograph No. MS-21). Princeton, NJ: ETS.

Stricker, L., & Attali, Y. (2010). [*Test takers' attitudes about the TOEFL iBT*](#) (TOEFL iBT™ Report No. iBT-13). Princeton, NJ: ETS.

Consequential Validity

Bailey, K. M. (1999). [*Washback in language testing*](#) (TOEFL® Monograph No. MS-15). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2006). [*The impact of changes in the TOEFL® examination on teaching and learning in Central and Eastern Europe: Phase 1, The baseline study*](#) (TOEFL® Monograph No. MS-34). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2008). [*The impact of changes in the TOEFL® examination on teaching and learning in Central and Eastern Europe: Phase 2, Coping with change*](#) (TOEFL iBT™ Report No. iBT-05). Princeton, NJ: ETS.

Wall, D., & Horák, T. (2011). [*The impact of changes in the TOEFL® examination on teaching and learning in Central and Eastern Europe: Phase 3, The role of the coursebook, and Phase 4, Describing change*](#) (TOEFL iBT™ Report No. iBT-17). Princeton, NJ: ETS.

Fairness and Accessibility

Candidate Characteristics

Breland, H., Lee, Y.-W., Najarian, M., & Muraki, E. (2004). [*An analysis of TOEFL® CBT writing prompt difficulty and comparability for different gender groups*](#) (TOEFL® Research Rep. No. RR-76). Princeton, NJ: ETS.

Lee, Y.-W., Breland, H., & Muraki, E. (2004). [*Comparability of TOEFL® CBT writing prompts for different native language groups*](#) (TOEFL® Research Rep. No. RR-77). Princeton, NJ: ETS.

Liu, L., Schedl, M., Malloy, J., & Kong, N. (2009). [*Does content knowledge affect TOEFL iBT™ reading performance? A confirmatory approach to differential item functioning*](#) (TOEFL iBT™ Report No. iBT-09). Princeton, NJ: ETS.

Stricker, L. J., Rock, D. A., & Lee, Y.-W. (2005). [*Factor structure of the LanguEdge™ test across language groups*](#) (TOEFL® Monograph No. MS-32). Princeton, NJ: ETS.

Stricker, L. J., & Rock, D. A. (2008). [*Factor structure of the TOEFL® Internet-based test across subgroups*](#) (TOEFL iBT™ Report No. iBT-07). Princeton, NJ: ETS.

Hill, Y. Z., & Liu, O. L. (2012). [*Is there any interaction between background knowledge and language proficiency that affects TOEFL iBT® reading performance?*](#) (TOEFL iBT™ Report No. iBT-18). Princeton, NJ: ETS.

Technology

Breland, H., Lee, Y.-W., & Muraki, E. (2004). [*Comparability of TOEFL® CBT writing prompts: Response mode analyses*](#) (TOEFL® Research Rep. No. RR-75). Princeton, NJ: ETS.

Hansen, E. G., Forer, D. C., & Lee, M. J. (2004). [*Toward accessible computer-based tests: Prototypes for visual and other disabilities*](#) (TOEFL® Research Rep. No. RR-78). Princeton, NJ: ETS.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). [*The relationship between computer familiarity and performance on computer-based TOEFL® test tasks*](#) (TOEFL® Research Rep. No. RR-61). Princeton, NJ: ETS.

Wolfe, E. W., & Manalo, J. R. (2005). [*An investigation of the impact of composition medium on the quality of scores from the TOEFL® writing section: A report from the broad-based study*](#) (TOEFL® Research Rep. No. RR-72). Princeton, NJ: ETS.

Support for Test Revision

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). [*TOEFL® 2000 listening framework: A working paper*](#) (TOEFL® Monograph No. MS-19). Princeton, NJ: ETS.

Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). [*TOEFL® 2000 speaking framework: A working paper*](#) (TOEFL® Monograph No. MS-20). Princeton, NJ: ETS.

Carrell, P. L. (2007). [*Notetaking strategies and their relationship to performance on listening comprehension and communicative assessment tasks*](#) (TOEFL® Monograph No. MS-35). Princeton, NJ: ETS.

Carrell, P. L., Dunkel, P. A., & Mollaun, P. (2002). [*The effects of notetaking, lecture length and topic on the listening component of the TOEFL® 2000 test*](#) (TOEFL® Monograph No. MS-23). Princeton, NJ: ETS.

Cumming, A., Kantor, R., Powers, D. E., Santos, T., & Taylor, C. (2000). [*TOEFL® 2000 writing framework: A working paper*](#) (TOEFL® Monograph No. MS-18). Princeton, NJ: ETS.

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. MS-17). Princeton, NJ: ETS.

Ginther, A. (2001). [*Effects of the presence and absence of visuals on performance on TOEFL® CBT listening-comprehension stimuli*](#) (TOEFL® Research Rep. No. RR-66). Princeton, NJ: ETS.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). [*TOEFL® 2000 framework: A working paper*](#) (TOEFL Monograph No. MS-16). Princeton, NJ: ETS.

Scoring and Technology

Scoring Issues

Brown, A., Iwashita, N., & McNamara, T. (2005). [*An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks*](#) (TOEFL® Monograph No. MS-29). Princeton, NJ: ETS.

Cumming, A., Kantor, R., & Powers, D. E. (2001). [*Scoring TOEFL® essays and TOEFL® 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework*](#) (TOEFL® Monograph No. MS-22). Princeton, NJ: ETS.

Erdosy, M. U. (2004). [*Exploring variability in judging writing ability in a second language: A study of four experienced raters of ESL compositions*](#) (TOEFL® Research Rep. No. RR-70). Princeton, NJ: ETS.

Jamieson, J., & Poonpon, K. (2013). [*Developing analytic rating guides for TOEFL iBT® integrated speaking tasks*](#) (TOEFL iBT Report No. iBT-20). Princeton, NJ: ETS.

Lee, Y.-W., Gentile, C., & Kantor, R. (2008). [*Analytic scoring of TOEFL® CBT essays: Scores from humans and e-rater®*](#) (TOEFL® Research Report No. RR-81). Princeton, NJ: ETS.

Lee, Y.-W., & Sawaki, Y. (in press). [*Psychometric Models for Cognitive Diagnosis of Assessments of English as a Second Language*](#) (TOEFL iBT™ Report No. iBT-12). Princeton, NJ: ETS.

Weigle, S.C. (2011). [*Validation of automated scores of TOEFL iBT® tasks against nontest indicators of writing ability*](#) (TOEFL iBT™ Report No. iBT-15). Princeton, NJ: ETS.

Winke, P., Gass, S., & Myford, C. (2011). [*The relationship between raters' prior language study and the evaluation of foreign language speech samples*](#) (TOEFL iBT™ Report No. iBT-16). Princeton, NJ: ETS.

Xi, X., & Mollaun, P. (2006). [*Investigating the utility of analytic scoring for the TOEFL® Academic Speaking Test \(TAST\)*](#) (TOEFL iBT™ Report No. iBT-01). Princeton, NJ: ETS.

Xi, X., & Mollaun, P. (2009). [*How do raters from India perform in scoring the TOEFL iBT™ speaking section and what kind of training helps?*](#) (TOEFL iBT™ Report No. iBT-11). Princeton, NJ: ETS.

Development of Automated Scoring Tools

Burstein, J. C., Kaplan, R. M., Rohen-Wolff, S., Zuckerman, D. I., & Lu, C. (1999). [*A review of computer-based speech technology for the TOEFL® 2000 test*](#) (TOEFL® Monograph No. MS-13). Princeton, NJ: ETS.

Chodorow, M., & Burstein, J. (2004). [*Beyond essay length: Evaluating e-rater®'s performance on TOEFL® essays*](#) (TOEFL® Research Rep. No. RR-73). Princeton, NJ: ETS.

Frase, L. T., Faletti, J., Ginther, A., & Grant, L. (1999). [*Computer analysis of the TOEFL® Test of Written English™*](#) (TOEFL Research Rep. No. RR-64). Princeton, NJ: ETS.

Leacock, C., & Chodorow, M. (2001). [*Automatic assessment of vocabulary usage without negative evidence*](#) (TOEFL® Research Rep. No. RR-67). Princeton, NJ: ETS.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). [*Automated scoring of spontaneous speech using SpeechRater v1.0*](#) (ETS Research Rep. No. RR-08-62). Princeton, NJ: ETS.

Zechner, K., Bejar, I. I., & Hemat, R. (2007). [*Towards an understanding of the role of speech recognition in non-native speech assessment*](#) (TOEFL iBT™ Report No. iBT-02). Princeton, NJ: ETS.

Candidates and Populations

Educational Testing Service (2014). [*Test and score data summary for TOEFL iBT® tests: January 2013-December 2013 test data*](#) (TOEFL®-SUM-13). Princeton, NJ: ETS.

Powell, W. W. (2001). [*Looking back, looking forward: Trends in intensive English program enrollments*](#) (TOEFL® Monograph No. MS-14). Princeton, NJ: ETS.

Reliability and Generalizability

Educational Testing Service. (2011). [*Reliability and comparability of TOEFL® iBT scores*](#) (TOEFL Research Insight Series Vol. 4). Princeton, NJ: ETS.

Lee, Y.-W. (2005). [*Dependability of scores for a new ESL speaking test: Evaluating prototype tasks*](#) (TOEFL® Monograph No. MS-28). Princeton, NJ: ETS.



Lee, Y.-W., & Kantor, R. (2005). [*Dependability of new ESL writing test scores: Evaluating prototype tasks and alternative rating schemes*](#) (TOEFL® Monograph No. MS-31). Princeton, NJ: ETS.

Zhang, Y. (2008). [*Repeater analyses for TOEFL® iBT*](#) (ETS Research Memorandum No. RM-08-05). Princeton, NJ: ETS.

Score Interpretation

Sawaki, Y., & Sinharay, S. (2013). [*Investigating the value of section scores for the TOEFL iBT® test*](#) (TOEFL iBT™ Report No. iBT-21). Princeton, NJ: ETS.

Stricker, L., & Wilder, G. (2012). [*Test takers' interpretation and use of TOEFL iBT® score reports: A focus group study*](#) (ETS Research Memorandum No. RM-12-08). Princeton, NJ: ETS.

Tannenbaum, R. J., & Wylie, E. C. (2008). [*Linking English-language test scores onto the common European framework of reference: An application of standard-setting methodology*](#) (TOEFL iBT™ Report No. iBT-05). Princeton, NJ: ETS.

Wylie, C. E., & Tannenbaum, R. J. (2006). [*TOEFL® Academic speaking test: Setting a cut score for international teaching assistants*](#) (ETS Research Memorandum No. RM-06-01). Princeton, NJ: ETS.

Xi, X. (2008). [*Investigating the criterion-related validity of the TOEFL® speaking scores for ITA screening and setting standards for ITAs*](#) (TOEFL iBT™ Report No. iBT-03). Princeton, NJ: ETS.

Please visit <http://www.ets.org/toefl/research> for the TOEFL® Research Reports.

Updated November, 2014

Copyright © 2014 by Educational Testing Service. All rights reserved. ETS, the ETS logo, LISTENING. LEARNING. LEADING., E-RATER and TOEFL are registered trademarks of Educational Testing Service (ETS). LANGUEDGE, TEST OF ENGLISH AS A FOREIGN LANGUAGE, TEST OF WRITTEN ENGLISH, and TOEFL iBT are trademarks of ETS. 12430