

TOEFL Junior® Research

Baron, P. A., & Tannenbaum, R. (2011). *Mapping the TOEFL Junior® Test onto the Common European Framework of Reference* (ETS Research Memorandum ETS RM-11-07). Princeton, NJ: Educational Testing Service.

This standard-setting study linked scores on the *TOEFL Junior* Standard test to the Common European Framework of Reference (CEFR). Fourteen language experts from nine countries served on the standard-setting panel and recommended to policymakers the minimum cut scores for each of the three CEFR levels on the three sections of the *TOEFL Junior* Standard test.

Evanini, K., Heilman, M., Wang, X., & Blanchard, D. (2015). *Automated Scoring for the TOEFL Junior® Comprehensive Writing and Speaking Test* (ETS Research Report No. RR-15-09). Princeton, NJ: Educational Testing Service.

This report describes the initial automated scoring results obtained using the constructed responses from the Writing and Speaking sections of the pilot forms of the *TOEFL Junior* Comprehensive test administered in 2011. Form-level results based on the five responses in the Writing section showed a human-machine correlation of $r = .83$, compared to a human-human correlation of $r = .90$. Form-level results based on the five items in the Speaking section showed a human-machine correlation of $r = .81$, compared to a human-human correlation of $r = .89$.

Gu, L. (2015). Language Ability of Young English Language Learners: Definition, Configuration, and Implications. *Language Testing*, 32(1), 21–38.

This study examines the dimensionality of the latent ability underlying language use that young learners need to function in English-medium instructional environments, where English is used as the means of instruction. Results showed that the two ability constructs (i.e., academic and social language), although theoretically distinct and educationally relevant, were statistically indistinguishable based on English as a foreign language (EFL) learners' performance on the *TOEFL Junior* Comprehensive test. Test performance could be explained best by a higher-order model, indicating that the language ability of these young EFL learners was structurally similar to that usually found with adult learners in a foreign language environment. The interpretation of young EFL learners' language proficiency needs to take into consideration how language components are developmentally related to each other as a function of learning experience in a foreign language environment.

Gu, L., Lockwood, J., & Powers, D. E. (2015). *Evaluating the TOEFL Junior® Standard Test as a Measure of Progress for Young English Language Learners* (ETS Research Report No. RR-15-22). Princeton, NJ: Educational Testing Service.

This study uses nonexperimental repeated measures data from approximately 4,600 students from multiple countries to examine the extent to which observed patterns in within-individual changes in test scores were consistent with changes in underlying language proficiency because of learning. The time interval between test administrations serves as a proxy for the extent of English language learning opportunities. Hierarchical linear models were used to model growth in test performance as a function of the time interval between test administrations. A positive, statistically significant relationship was found between score gain and the length of the time interval between testing and retesting: test takers with longer intervals between test administrations showed greater gains than did test takers who retested at shorter intervals. Depending on model specification, the estimated relationship for the total score corresponded to between .16 and .24 standard deviations of growth per year. These findings suggest that the *TOEFL Junior* standard test is capable of reflecting change in English-language proficiency over time.

Papageorgiou, S., & Cho, Y. (2014). An Investigation of the Use of *TOEFL Junior®* Standard Scores for ESL Placement Decisions in Secondary Education. *Language Testing*, 31(2), 223–239.

This study examined the relationship between secondary school students' *TOEFL Junior* Standard test scores and the placement of these students into ESL classes, and found strong correlations between test scores and the teacher-assigned ESL levels. The findings provided some preliminary evidence to support the use of the *TOEFL Junior* Standard test as an initial screening tool for ESL placement.

Papageorgiou, S., Morgan, R., & Becker, V. (2015). Enhancing the Interpretability of the Overall Results of an International Test of English-Language Proficiency. *International Journal of Testing*, 15(4), 310–336.

The purpose of this study was to develop performance levels and descriptors to accompany the total scale scores of *TOEFL Junior* Standard tests. Similar to an earlier study for the *TOEFL Junior* Comprehensive test, the study addressed two issues: the number of performance levels that could be meaningfully reported and the information that should be included in the performance level descriptors. Data from 3,607 students who took an operational test form were used. Although our methodology built to some extent on the earlier *TOEFL Junior* Comprehensive study, we demonstrate how content and construct differences between the tests of each study dictated use of different types of data in order to construct meaningful performance descriptors and select the cut-offs for the levels.

Papageorgiou, S., Xi, X., Morgan, R., & So, Y. (2015). Developing and Validating Band Levels for Reporting Overall Examinee Performance. *Language Assessment Quarterly*, 12(2), 153–177.

This study presents the development and empirical validation of score levels and descriptors specifically designed for reporting purposes in the context of the *TOEFL Junior* Comprehensive test. Test performance data from 2,931 students were used. The band level solution was determined by balancing considerations for the reliability of classification decisions and the desire for the levels to represent meaningful performance differences. Meaningful descriptors for the band levels were constructed using the scoring rubrics, the characteristics of test items, typical student performance profiles and the performance of norm groups on the test.

So, Y. (2014). Are Teacher Perspectives Useful? Incorporating EFL Teacher Feedback in the Development of a Large-Scale International English Test. *Language Assessment Quarterly*, 11(3), 283–303.

This case study shows how English teachers' perspectives were incorporated into the development of a large-scale international English assessment, the *TOEFL Junior* Comprehensive test. The article discusses how stakeholder feedback gathered during test development supports the validity argument for score interpretation and the use of a newly developed test. When the pilot version of the test was administered, focus-group interviews were conducted with 10 teachers of English as a foreign language (EFL) in Korea. The article summarizes the teachers' major comments, recommendations and suggestions about the test and shows how these were reflected in the final test design decisions.

So, Y., Wolf, M. K., Hauck, M. C., Mollaun, P., Rybinski, P., Tumposky, D., & Wang, L. (2015). *TOEFL Junior® Design Framework (TOEFL® Young Students Research Report No. TOEFL Jr-02)*. Princeton, NJ: Educational Testing Service.

This paper presents the theoretical and empirical foundations of the *TOEFL Junior* assessment and its development process. The *TOEFL Junior* test was developed to address the increasing need for objective measures of English-language proficiency for young adolescent learners, who are being introduced to English as a second or foreign language at a much younger age than ever before. This report presents the test purposes and intended uses, target population, target language use domains, and test constructs of the *TOEFL Junior* test, along with a description of the overall test structure and scoring system, which demonstrates how the constructs are operationalized. Finally, we outline research topics to support the interpretive argument of the use of the test. This document is expected to serve as a reference point during investigations of validity evidence to support the intended test uses over time.

Wolf, M.K., & Steinberg, J. (2011). *An Examination of United States Middle School Students' Performance on TOEFL Junior®*. ETS Research Memorandum (ETS-RM-11-15). Princeton, NJ: Educational Testing Service.

This study examined how native English-speaking and English-language learner students in U.S. middle schools perform on the *TOEFL Junior* Standard test. The percentile rank and statistical significance test results indicated that there was a clear pattern of increasing student performance on the *TOEFL Junior* Standard test from grades 6–8.

Young, J. W., Morgan, R., Rybinski, P., Steinberg, J., & Wang, Y. (2013). *Assessing the Test Information Function and Differential Item Functioning for the TOEFL Junior® Standard Test*. (TOEFL Young Learners Research Report No. TOEFL-YL-01). Princeton, NJ: Educational Testing Service.

This study investigated two psychometric characteristics of the *TOEFL Junior* Standard test: (1) the use of test scores in placement decisions and (2) the level of Differential Item Functioning (DIF) across test takers from different countries. Results indicated sufficient information to support placement decisions across most score scales for all three assessment sections, and although some items exhibited significant DIF, further analyses showed plausible construct-relevant explanations for most of the findings.