TOEFL Junior® Research

Summary of TOEFL Junior Research Flyer

The studies featured and summarized on this research flyer provide a cumulative body of validity and fairness evidence for the TOEFL Junior® tests. Consistent with the *ETS Standards for Quality and Fairness*, fairness is treated as an integral part of validity. For instance, score interpretations should be appropriate for the intended young adolescent test-taker population, and test performance should not be significantly influenced by construct-irrelevant factors such as developmental stage, educational context, language-learning background, or mode of test delivery. Across studies, the evidence supports the appropriateness of TOEFL Junior for assessing English proficiency among young learners in diverse international settings. Several studies examine developmental and age-related considerations, including grade-level differences, working memory, and young students' test-taking strategies (Brunfaut et al., 2021; Choi & Loewen, 2022; Wolf & Steinberg, 2011). Other studies address fairness and validity across linguistic, national, and educational contexts by examining stakeholder perceptions, curricular alignment, local score interpretation, and progress monitoring in settings such as Armenia, Austria, China, Germany, South Korea, Taiwan, and other countries (Galikyan et al., 2019; Hsieh, 2024; Huang et al., 2020; Kormos et al., 2024; Madyarov et al., 2021; Papageorgiou et al., 2022; Timpe-Laughlin, 2018). Findings generally indicate that the TOEFL Junior tests are developmentally appropriate, aligned with intended academic English constructs, adequate for placement and progress-monitoring decisions, and interpretable across diverse learner populations and educational contexts. Together, this evidence supports the use of TOEFL Junior scores for their intended purposes while underscoring the importance of continued monitoring of subgroup performance, accessibility, and local score use in accordance with ETS fairness standards.

Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners' reading comprehension: Exploring the effects of working memory, grade level, and reading task. Language Testing, 38 (1), 356-377. https://doi.org/10.1177/0265532221991480
This study investigates the effects of working memory (WM), grade level, and reading task on young learners' foreign language reading test performances. Ninety-four young English language learners (Grades 6–7) in Hungary completed the TOEFL Junior comprehensive's reading test and a WM test battery. The mixed-effects model predicted significantly higher comprehension accuracy among learners with higher WM capacity, and among learners in Grade 7 compared to learners in Grade 6. Reading task differences were not associated with significant comprehension accuracy differences. The findings emphasize the importance of cognitive and strategic factors in reading comprehension and highlight implications for designing developmentally appropriate reading assessments.

Choi, J. S., & Loewen, S. (2022). Exploring young learners' strategic behaviors in a speaking test. TESOL Quarterly, 56(4), 1384–1396. https://doi.org/10.1002/tesq.3136
This study investigated the strategic behaviors of 45 young learners (ages 11–14) in South Korea while completing TOEFL Junior Speaking practice tasks. Using stimulated recall protocols, the study identified 40 strategies across five categories, with communication and metacognitive strategies most frequent. Learners' strategy use varied by task type (picture narration vs. listen-speak), and moderate correlations emerged between certain strategies (e.g., rehearsing, pre-task planning) and higher performance scores. Findings highlight task-specific and age-appropriate aspects of strategy use, underscoring the role of strategic competence in validating speaking assessments for young learners.

Galikyan, G., Madyarov, I., & Gasparyan, R. (2019). Student test-takers' and teachers' perceptions of the TOEFL Junior standard test (TOEFL Research Report RR-19-29) Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12264
This study examined perceptions of the TOEFL Junior standard test among 202 adolescent EFL learners and 9 teachers in an afterschool program in Armenia. Using questionnaire data, the study explored stakeholder views on the test's appropriateness and effectiveness. Results indicated that students found the tasks developmentally suitable and reflective of their English abilities, while teachers valued the test's ability to provide meaningful proficiency information. The findings support the test's construct validity, offering evidence that test content aligns with users' expectations and
learning goals. Findings reinforce TOEFL Junior's appropriateness for use in diverse EFL educational settings.

Hsieh, C.-N. (2024). Building a validity argument for the TOEFL Junior tests (TOEFL Research Report No. RR-102; ETS Research Report Series No. RR-24-05). Educational Testing Service. https://doi.org/10.1002/ets2.12379
This research report advances the validity argument for the TOEFL Junior assessments, designed to measure English language proficiency among secondary-level learners in English-medium contexts. The report structures the validity argument through six inferential steps—domain description, evaluation, generalization, explanation, extrapolation, and utilization—each supported by empirical evidence. Drawing on psychometric analyses, rater reliability studies, curricular alignment research, and correlations with teacher judgments, self-assessments, and external proficiency measures, the report demonstrates that TOEFL Junior scores provide stable, interpretable, and theoretically grounded indicators of academic English proficiency. The findings substantiate the test's appropriateness for placement, progress monitoring, and instructional decision-making.

Hsieh, C.-N. (2024). The role of task types and reading proficiency on young English as a foreign language learners' writing performances. TESOL Quarterly, 58(2), 978–990. https://doi.org/10.1002/tesq.3286
This study explored how writing task type and L2 reading proficiency affect young EFL learners' writing performance. A total of 185 Grade 7–8 students from Denmark, Finland, and the Netherlands completed a descriptive and an email writing task. Responses were scored holistically and analyzed for lexical sophistication, syntactic complexity, fluency, and idea development. Students were grouped as strong or emergent readers based on TOEFL Junior reading scores. Descriptive tasks elicited more complex language, while strong readers outperformed peers in fluency and idea development. Results underscore the link between reading and writing, highlighting the importance of strengthening reading to support writing development.

Huang, B. H., Bailey, A. B., Sass, D., & Chang, Y. S. (2020). An investigation of the validity of a speaking assessment for adolescent English language learners. Language Testing, 38(3) 401–428. https://doi.org/10.1177/0265532220925731 This study investigated the construct validity of the TOEFL Junior Speaking Test for adolescent EFL learners in Taiwan, using Messick's and Kane's validity frameworks. A total of 252 seventh-to-ninth graders took the TOEFL Junior Speaking Test and two additional researcher-developed speaking tasks. Students and teachers also evaluated English language skills and learning histories. Confirmatory factor analysis provided strong support for the test's internal structure. Correlations and regression analyses demonstrated significant positive relationships between TOEFL Junior scores and both self- and teacher-reported language abilities. The findings offer robust evidence for the test's validity and appropriateness for assessing adolescent EFL speaking proficiency.

Kormos, J., Eberharter, K., Guggenbichler, E., Baumgartinger, S., Ebner, V., & Kremmel, B. (2024). First Language Literacy Skills, Listening Strategy Use, and Anxiety in Self-Paced versus Single-Play L2 Listening Tests. Language Assessment Quarterly, 21(4–5), 303–323. https://doi.org/10.1080/15434303.2024.2435811
This study explored how single play versus self-paced listening test formats affect young Austrian EFL learners' listening anxiety, strategy use, and performance. Using the TOEFL Junior standard test's listening section, students completed both listening modes and answered questions on strategy use and anxiety. They also took a standardized L1 literacy test. Students reported lower anxiety in the self-paced condition, but strategy use remained largely unchanged across modes, except for greater concentration in the single-play format. No significant relationship was found between the frequency of self-pacing use and listening scores. Results suggest that self-pacing may not alter the construct measured by L2 listening tests.

Madyarov, I., Movsisyan, V., Madoyan, H., Galikyan, I. and Gasparyan, R. (2021). New validity evidence on the TOEFL Junior standard test as a measure of progress (ETS Research Report No. 21). ETS. https://doi.org/10.1002/ets2.12334 This study examined the TOEFL Junior standard test as a measure of English language progress among adolescents in a controlled after-school program in Armenia. A total of 154 learners took the test three times over 30 instructional weeks. Results from a difference-in-differences (DID) analysis showed that the test was sensitive to progress among A1–A2 learners after just 20 hours of instruction per 10 weeks. However, no significant gains were detected for B1–B2 learners despite double the instructional time. The findings support the test's use as a progress indicator for lower-level learners.

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. https://doi.org/10.1177/0265532213499750 This study examined the relationship between secondary school students' TOEFL Junior standard test scores and the placement of these students into English as a Second Language (ESL) classes and found strong correlations between test scores and the teacher-assigned ESL levels. The findings provided preliminary evidence to support the use of the TOEFL Junior standard test as an initial screening tool for ESL placement.

Papageorgiou, S., Wu, S., Hsieh, C-N., Tannenbaum, R. J., Cheng, M. (2022). Aligning language test scores to local proficiency levels: The case of China's Standards of English Language Ability (CSE). Chinese/English Journal of Educational Measurement and Evaluation, 3(1), 1-18. https://www.ce-jeme.org/journal/vol3/iss1/1 This paper explores the process of aligning test scores from the TOEFL Junior Standard test and the TOEFL Junior Speaking test to China's Standards of English Language Ability (CSE), a localized framework for English as a foreign language. The authors detail methodological steps for alignment, emphasizing the need for established procedures, documentation, and multiple sources of evidence. They also highlight contextual factors influencing score interpretation and the high-stakes implications for individuals and institutions. The study provides a model for future alignment research, especially when adapting global assessments to local proficiency standards.

So, Y., Wolf, M. K., Hauck, M. C., Mollaun, P., Rybinski, P., Tumposky, D., & Wang, L. (2015). TOEFL Junior design framework (TOEFL Junior Research Report TOEFL JR–02). Princeton, NJ: Educational Testing Service. https://files.eric.ed.gov/fulltext/EJ1109688.pdf This paper presents the theoretical and empirical foundations of the TOEFL Junior assessments 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 learning English as a second or foreign language. This report presents the test purposes and intended uses, target population, target language use domains, and test constructs of the TOEFL Junior tests along with a description of the overall test structure and scoring system, which demonstrates how the constructs are operationalized. Finally, it outlines research topics to support the interpretive argument of the use of the test. This document has been used as a reference point during investigations of validity evidence to support the intended test uses over time.

Timpe-Laughlin, V. (2018). A good fit?— Examining the alignment between the TOEFL Junior standard test and the English as a foreign language curriculum in Berlin, Germany. (Research Memorandum No. RM-18-11). Princeton, NJ: Educational Testing Service. https://www.ets.org/Media/Research/pdf/RM-18-11.pdf This case study examined whether the TOEFL Junior standard test aligns with Berlin's secondary-level EFL curriculum, given growing interest in using it as a *Klassenarbeitsersatz* (substitute for a mandatory classroom assessment). Researchers reviewed Berlin's official curriculum and systematically coded activities from textbooks used in Grades 7–10. Findings showed substantial overlap between test content and instructional materials. Triangulation with teacher feedback further supported the test's curricular relevance. Results suggest that the TOEFL Junior standard test is a valid tool for assessing EFL proficiency in this local context and can appropriately serve as an external measure within schools in Berlin.

Wolf, M. K., & Steinberg, J. (2011). An examination of United States middle school students' performance on TOEFL Junior (ETS Research Memorandum No. ETS-RM-11-15). Educational Testing Service. http://www.ets.org/Media/Research/pdf/RM-11-15.pdf 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.

Wolf, M. K., Suhan, M., & Alderman, J. (2024, April). Exploring the provision of artificial intelligence–based feedback for TOEFL Junior writing practice tasks (ETS Research Memorandum No. RM–24-05). Educational Testing Service. **http://www.ets.org/Media/Research/pdf/RM-24-05.pdf** This research memorandum reports on an exploratory study of developing an artificial intelligence (AI)–based feedback prototype tool and examining its usability for potential TOEFL Junior Writing test users. The TOEFL Junior Writing test employs ETS's automated writing evaluation (AWE) tool to score test takers' responses. Utilizing ETS's AWE engine and an existing database, the authors designed an approach to feedback that was closely linked to the key writing dimensions and proficiency descriptors as outlined in the TOEFL Junior Writing rubrics. A small-scale usability study was conducted with 14 students and seven teachers from South Korea and Türkiye. The participants perceived the usefulness of the feedback tool positively, while also providing practical suggestions to further improve the tool. The promises and limitations of the tool are discussed.

Copyright © 2026 by ETS. ETS, the Eight-Point logo, and TOEFL Junior are registered trademarks of ETS in the United States and other countries."