

# TOEFL Primary<sup>®</sup> Research

**Cho, Y., & Blood, I. (2020). An analysis of TOEFL Primary repeaters: How much score change occurs? *Language Testing*, 37(4), 503-522. <https://doi.org/10.1177/0265532220927751>**

This study investigated expected score changes on the TOEFL Primary Listening and Reading tests based on the time interval between administrations. Data from 5,213 young learners (ages 8–13) in Japan and Turkey were analyzed using multilevel modeling to account for individual differences in initial scores and growth rates. The study also examined the influence of age and test-level difficulty. Analyses by country revealed a gradual increase in scores over time, with growth rates differing between Japan and Turkey. Results suggest that older students and longer intervals between testings are associated with higher gains, offering guidance for re-testing decisions.

**Cho, Y., & So, Y. (2014). Construct-Irrelevant Factors Influencing Young English as a Foreign Language (EFL) Learners' Perceptions of Test Task Difficulty (ETS Research Memorandum ETS RM 14-04). ETS. <https://web.archive.org/web/20210721122925/https://www.ets.org/s/research/pdf/RM-14-04.pdf>**

The objective of the study was to understand the factors that may unduly influence students' performance on an English test. Twelve Korean primary school children (ages 9–11) who were learning English as a foreign language (EFL) interacted with test materials and responded to English language test questions. Participants were also interviewed to elicit their perception of the difficulty of test questions and to gain insight into their thought processes while they were answering listening and reading test questions. Verbal reports indicated that the complexity of language used in test questions may interfere with students' demonstration of their abilities on the constructs targeted in the test questions. Memory load and ambiguity in task descriptions were also observed to influence young test takers' perception of test difficulty.

**Cho, Y., Ginsburgh, M., Morgan, R., Moulder, B., Xi, X., & Hauck, M. C. (2016). Designing the TOEFL<sup>®</sup> Primary<sup>™</sup> tests (ETS Research Memorandum No. RM-16-02). ETS. <https://www.ets.org/Media/Research/pdf/RM-16-02.pdf>**

This research memorandum outlines the conceptual framework and design principles guiding the development of the TOEFL Primary tests, which assess English language proficiency in young learners aged 8+. Grounded in a socio-cognitive validity framework, the test design aligns with real-world language use in social and academic contexts. The report details construct definition, domain specification, task design, and scoring, emphasizing age-appropriate content and cognitive demands. It also highlights the interpretive argument and intended test uses. The memorandum serves as a foundation for ongoing validation research and supports stakeholders in interpreting test scores for instructional and evaluative purposes.

**Hsieh, C. (2016). Examining Content Representativeness of a Young Learner Language Assessment: EFL Teachers' Perspectives. In M. Nikolov (Ed.), Assessing Young Learners of English: Global and Local Perspectives. (93–108). Springer.**

This study examines the content validity of the TOEFL Primary test by employing expert judgments. Seventeen experienced EFL teachers from 15 countries participated in the study. The teachers evaluated the relevance and importance of the language knowledge and skills the TOEFL Primary test is intended to measure. The expert teachers' judgments were used as the criterion on which the content-related evidence of validity was based. Results of the study indicated that the TOEFL Primary test content measures what it is intended to measure and covers the important domains of language knowledge and skills young language learners are required to possess to perform effectively in EFL classrooms.

**Hsieh, C.-N. (2024). Building a validity argument for the TOEFL Primary tests (TOEFL Research Report No. RR-103 & ETS Research Report No. RR-24-16). ETS. <https://www.ets.org/Media/Research/pdf/RR-24-16.pdf>**

The TOEFL Primary tests measure young learners' English language knowledge and skills for communication in social and academic contexts at the primary and lower secondary levels. This paper presents a comprehensive validity argument, comprising six inferences and a network of assumptions about score use and measurement quality. It presents a considerable body of evidence that supports the TOEFL Primary tests as valid tools for assessing proficiency, monitoring progress, and informing instruction. The paper offers guidance to score users and stakeholders on interpreting results and making informed decisions about the test's suitability for various educational contexts.

**Hui, B., Wong, S. S. Y., & Au, R. K. C. (2022). Reading aloud listening test items to young learners: Attention, item understanding, and test performance. *System*, 108, 102831. <https://doi.org/10.1016/j.system.2022.102831>**

In young learner L2 listening assessment, test items are often read aloud, yet empirical support for this practice remains limited. This study investigates how read-alouds affect attention, perceived item understanding, and performance. Non-native English-speaking children completed a baseline reading test, then took a sample TOEFL Primary listening test under two conditions—fully read aloud (FRA) and partially read aloud (PRA)—while eye movements were tracked. After each item, children rated their understanding. Read-alouds enhanced visual attention to text but did not significantly affect perceived understanding or performance. Implications for assessment design and administration are discussed.

**Lee, S., & Winke, P. (2018). Young learners' response processes when taking computerized tasks for speaking assessment. *Language Testing*, 35(2), 239-269. <https://doi.org/10.1177/0265532217704009>**

This study examined how young learners process and perceive a computer-based, timed speaking test. Twenty non-native (NNS) and eight native English-speaking (NS) children (ages 8–10) completed seven TOEFL Primary speaking tasks. Eye-tracking data revealed differences in attentional focus: NNSs fixated more often and longer on the countdown timer, while NSs focused more on content-related visuals. Timer fixations among NNSs frequently coincided with hesitation behaviors. Performance scores also differed significantly between groups. Interviews and drawings supported these findings. Results highlight how test design features may impact young learners' performance and inform child-appropriate testing accommodations and practices.

**Papageorgiou, S., & Baron, P. A. (2017). Using the Common European Framework of Reference for young learners' English language proficiency assessments. In M. K. Wolf & Y. G. Butler (Eds.), English language proficiency assessments for young learners (pp. 136-152). Routledge.**

The chapter explored how the Common European Framework of Reference for Languages (CEFR) can inform the development and interpretation of English language assessments for young learners. It outlines how the CEFR has been used to inform the development and score interpretation of the TOEFL Young Student tests, including the TOEFL Primary tests. The authors highlight how CEFR-aligned proficiency bands help situate young learners' performance within an internationally recognized framework, supporting global score interpretation. They discuss the challenges of aligning CEFR descriptors—designed for older learners—with younger test takers and emphasize the importance of age-appropriate test design and validation.

**Papageorgiou, S., & Choi, I. (2018). Adding value to second-language listening and reading subscores: Using a score augmentation approach. International Journal of Testing, 18(3), 207-230. <https://doi.org/10.1080/15305058.2017.1407766>**

This study investigated whether reporting subscores for item groups within reading and listening sections of second-language tests provides meaningful additional information beyond total section scores. Using data from 116,489 test takers across two large-scale English as a foreign language assessments, the authors applied Haberman's (2008) score augmentation method to enhance subscore reliability. Results showed limited psychometric support for reporting subscores, with only marginal gains from augmentation. Findings suggest that subscores add little measurement value.

**Papageorgiou, S., Xu, X., Timpe-Laughlin, V., & Dugdale, D. M. (2020). Exploring the alignment between a curriculum and a test for young learners of English as a foreign language. (Research Memorandum No. RM-20-08). ETS. <https://www.ets.org/Media/Research/pdf/RM-20-08.pdf>**

This study investigates the suitability of the TOEFL Primary tests for assessing English language proficiency in students enrolled in the online VIPKid Major Course (MC), Levels 3–7. Using test score data from the TOEFL Primary Listening and Reading tests and expert alignment judgments, the study found the tests to be generally appropriate in difficulty for MC students. Results also indicated increasing proficiency across course levels. Content alignment between the tests and MC curriculum was strong. Implications are discussed regarding the use of EFL assessments for young learners broadly, and the TOEFL Primary tests within the VIPKid MC context specifically.

**Suhan, M., Papageorgiou, S., & Wolf, M. K. (2024). Mapping the scores of the TOEFL Primary Writing test to the Common European Framework of Reference (CEFR) levels (Research Memorandum ETS RM-24-03). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-03.pdf>**

This research memorandum reports on a study to map the scores of the TOEFL Primary Writing test to the language proficiency levels of the Common European Framework of Reference for Languages (CEFR). As the TOEFL Primary Writing test is used globally, mapping its scores to the levels of the CEFR helps relevant stakeholders to interpret test results in relation to a globally used language framework as a point of reference. This report details the process undertaken, utilizing the judgment of expert panelists to establish recommended minimum test scores (cut scores) to classify test takers into CEFR levels.

**Timpe-Laughlin, V. (2015). Evaluating a Learning Tool for Young English Learners: The Case of the TOEFL® Primary™ English Learning Center (Research Report No. RR-15-04). ETS. <https://www.ets.org/Media/Research/pdf/RR-15-04.pdf>**

This report examines the use and perceptions of the English Learning Center (ELC), an online platform designed to support young English language learners' communication skills. Through surveys and interviews with teachers and elementary students, the study found that the ELC was primarily used as a supplementary, self-access tool outside EFL classrooms and was valued for providing language practice and TOEFL Primary test preparation. Findings offered preliminary support for claims made by ETS and Edusoft regarding the ELC's effectiveness. The report concludes with recommendations for improving the ELC and highlights considerations for future materials evaluation research with young learners.

**Wolf, M. K., Suhan, M., Ginsburgh, M., Futagi, Y., & Li, F. (2024). Design framework for the TOEFL Primary Writing test (Research Memorandum No. RM-24-02). ETS. <https://www.ets.org/Media/Research/pdf/RM-24-02.pdf>**

The TOEFL Primary tests, used internationally since 2013, assess English proficiency in young learners aged 8 and older. Initially focused on reading, listening, and speaking, the 2023 introduction of the TOEFL Primary Writing test enabled comprehensive four-skill assessment. This research memorandum outlines the theoretical design principles informing the development of the writing test and provides technical details about its structure and use. It also delineates directions for future validity research to support appropriate interpretation and use of TOEFL Primary Writing scores within global English language learning contexts.