TOEFL® Research Report

TOEFL–RR-88

ETS Research Report No. RR–19-29

Student Test Takers' and Teachers' Perceptions of the *TOEFL Junior*® Standard Test

Irena Galikyan

Irshat Madyarov

Rubina Gasparyan

December 2019

The *TOEFL*® test is the world's most widely respected English language assessment, used for admissions purposes in more than 130 countries including Australia, Canada, New Zealand, the United Kingdom, and the United States. 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, 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, the TOEFL Family of Assessments has expanded to provide high-quality English proficiency assessments for a variety of academic uses and contexts. The TOEFL Young Students Series (YSS) features the *TOEFL*® *Primary*™ and *TOEFL Junior*® tests, designed to help teachers and learners of English in school settings. The *TOEFL ITP*® Assessment Series offers colleges, universities, and others an affordable test for placement and progress monitoring within English programs.

Since the 1970s, the TOEFL tests have had a rigorous, productive, and far-ranging research program. ETS has made the establishment of a strong research base a consistent feature of the development and evolution of the TOEFL tests, because only through a rigorous program of research can a testing company demonstrate its forward-looking vision and substantiate claims about what test takers know or can do based on their test scores. In addition to the 20-30 TOEFL-related research projects conducted by ETS Research & Development staff each year, the TOEFL Committee of Examiners (COE), composed of distinguished language-learning and testing experts from the academic community, funds an annual program of research supporting the TOEFL family of assessments, including projects carried out by external researchers from all over the world.

To date, hundreds of studies on the TOEFL tests have been published in refereed academic journals and books. In addition, more than 300 peer-reviewed reports about TOEFL research have been published by ETS. These publications have appeared in several different series historically: TOEFL Monographs, TOEFL Technical Reports, TOEFL iBT Research Reports, and TOEFL Junior Research Reports. It is the purpose of the current TOEFL Research Report Series to serve as the primary venue for all ETS publications on research conducted in relation to all members of the TOEFL Family of Assessments.

Current (2019–2020) members of the TOEFL COE are:

Lia Plakans – Chair	**The University of Iowa**
Beverly Baker	University of Ottawa
April Ginther	Purdue University
Claudia Harsch	University of Bremen
Lianzhen He	Zhejiang University
Volker Hegelheimer	Iowa State University
Gerriet Janssen	Universidad de los Andes - Colombia
Lorena Llosa	New York University
Carmen Muñoz	The University of Barcelona
Yasuyo Sawaki	Waseda University
Randy Thrasher	International Christian University
Dina Tsagari	Oslo Metropolitan University

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org   **Web site:** www.ets.org/toefl

Student Test Takers' and Teachers' Perceptions of the *TOEFL Junior*® Standard Test

Irena Galikyan, Irshat Madyarov, & Rubina Gasparyan

Center for Research in Applied Linguistics, College of Humanities and Social Sciences, American University of Armenia, Yerevan, Republic of Armenia

The broad range of English language teaching and learning contexts present in the world today necessitates high quality assessment instruments that can provide reliable and meaningful information about learners' English proficiency levels to relevant stakeholders. The *TOEFL Junior*® tests were recently introduced by Educational Testing Service (ETS) to address the assessment needs of educators of young adolescent learners by providing necessary information on their English language proficiency. This study explores the perceptions of key stakeholders (202 student test takers and 9 teachers) of the TOEFL Junior Standard test in an English as a foreign language (EFL) context (an afterschool program in Armenia). The analysis of the data gathered through questionnaires provides insights into the perceptions of test users, suggesting that the test tasks were perceived to be developmentally appropriate for the student test takers and allowed them to demonstrate their English language abilities. The findings of the current study can serve as additional validity evidence for the TOEFL Junior Standard as they reflect the correspondence found between the test construct and test users' perceptions and, thus, support the validity argument for TOEFL Junior Standard score interpretation and use in EFL settings.

Keywords The *TOEFL Junior*® Standard test; stakeholder perceptions; construct-related validity

The global spread of English has resulted in a considerable increase in the number of young adolescent learners of English around the world, generating a growing need for the assessment of the English language proficiency of school-age learners in different settings (Wolf & Butler, 2017). This growth has stimulated the demand for high-quality standardized assessment instruments to measure young adolescent language learners' proficiency level in English and to monitor their progress over time. An example of such assessment is the *TOEFL Junior*® Standard test launched in 2010 by Educational Testing Service (ETS). The TOEFL Junior Standard test scores are intended to provide information for supporting decisions regarding placement of students into different instructional levels and monitoring student progress in developing English language proficiency over time (So et al., 2017).

Every assessment instrument is required to have validity evidence that supports score interpretation and use (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). However, the ongoing diversification of the target population of English learners poses challenges for assessment instruments (Butler, 2017). Thus, questions such as "how the stakeholders use the exams and what they think about them" (Saville, 2003, p. 60, Saville, 2010) may affect the validity of such large-scale tests as TOEFL Junior Standard in different settings. Therefore, as part of the test validation process, evaluating stakeholders' perceptions and attitudes toward a test is crucial in providing validity evidence on test interpretation and use (Cheng & DeLuca, 2011; Moss, Girard, & Haniford, 2006). The objective of this study is to evaluate the perceptions of two key stakeholder groups (student test takers and teachers) of the TOEFL Junior Standard test as a tool for assessing young learners' English language proficiency.

Background

TOEFL Junior Standard was recently introduced by ETS to address the assessment needs of educators of young adolescent learners between primarily 11 and 15 years of age. TOEFL Junior Standard is a paper-based, large-scale proficiency test (available digitally since 2018) designed to measure English language proficiency with respect to the academic and social English language skills in both English as a foreign language (EFL) and English as a second language (ESL) contexts.

Corresponding author: I. Galikyan, E-mail: igalikyan@aua.am

The test is composed of selected-response tasks in three sections: listening comprehension (Listening), language form and meaning (Language), and reading comprehension (Reading). The section items are gap-filling tasks that measure test takers' ability to (a) recognize a proper grammatical structure within context and (b) identify an appropriate lexical item within context (So et al., 2017).

The test lasts for 1 hour and 55 minutes. Each section has 42 items, making a total of 126 items. Section scores are provided on a scale from 200 to 300 points; the sum of the three section scores makes up the total score, ranging from 600 to 900. The TOEFL Junior Standard score report contains (a) an overall score level; (b) subscores for each of the sections; (c) a Common European Framework of Reference for Languages (CEFR) level for each test section, ranging from below A2 (the lowest level measured by the test), to A2, B1, and B2 (the highest level measured by the test); and (d) can-do statements describing English language abilities typical of the given CEFR level. The TOEFL Junior Standard test scores are intended to provide information for supporting decisions regarding placement of students into different instructional levels and monitoring student progress in developing English language proficiency over time. Some literature supports the practicality and financial benefit of using such large-scale tests over locally developed tests (e.g., Brown, 2004; Green, 2012).

The information obtained through tests and upon which decisions are made should be reliable and valid (Bachman, 1990). There is some empirical evidence supporting the validity of TOEFL Junior Standard, including criterion-related validity and construct validity based on the correlation between TOEFL Junior Standard scores and teacher judgments of proficiency and learners' TOEFL Junior Standard score gains over time (e.g., Gu, Lockwood, & Powers, 2015; Papageorgiou & Cho, 2014; Wolf & Steinberg, 2011).

In present-day language assessment, test validation entails utilization of various frameworks, multiple procedures, and multiple perspectives reflecting the growing emphasis put on the contextual factors and social consequences of a test (McNamara, 2007) and the importance of stakeholder voices in establishing test validity (Cheng & DeLuca, 2011). Stakeholders are directly impacted by a test through their interaction with its "construct, format, conditions and assessment criteria in various ways" (Hawkey, 2006, p. ix). These interactions determine the various beliefs and/or anxieties about the features of an assessment instrument that different stakeholders may have (Gu & So, 2015). When a discrepancy exists between the beliefs of test takers and test developers about what is actually measured in a test (e.g., Shohamy, Donitsa-Schmidt, & Ferman, 1996), it calls for more research to establish test validity in a particular context (Bachman, 2005, p. 32).

The inherent impact that affective schemata (e.g., motivation, attitudes, self-esteem) have on assessment performance (Bachman & Palmer, 2010) can be stronger with young learners (Wolf & Butler, 2017). In fact, research has shown that young learners' perceptions of tests can have considerable effect on their test performance, with negative perceptions leading to negative effects on performance (e.g., Aydin, 2012). The impact of factors such as task-related characteristics, test length, and test-taking experience on test performance suggests that assessment task demands should be consistent with the cognitive abilities of young learners (Aydin, 2012; Cho & So, 2014; Hasselgreen, 2005; McKay, 2006). This further emphasizes the need for investigating how young learner characteristics may interact with the features of a particular assessment instrument in different contexts (Butler, 2017). In addition, young learner assessment involves the complex dynamics of the relationships between multiple stakeholders, among which are teachers (Chik & Besser, 2011; Hasselgreen, 2005) whose "expert judgments and opinions can be viewed as part of a test's validity argument" (Winke, 2011, p. 628). This dynamic makes it crucial to investigate teachers' understanding of an assessment instrument—their perceptions and attitudes toward it—in order to discern a link between test validation and test use (Malone & Montee, 2014; Wall, 2000).

Thus, it is important to examine young learners' and teachers' perceptions of the TOEFL Junior Standard test in order to collect sufficient validity evidence "on how any assessment is situated in the local context" (Moss et al., 2006, p. 124) and on the meaningfulness and appropriateness of its score interpretations in that context (Messick, 1990). The main goal of this study was to explore student test takers' and teachers' perceptions of the TOEFL Junior Standard test in an EFL context. To this end, the following research questions guided the current study:

1. What are students' and teachers' perceptions of what the TOEFL Junior Standard test measures?
2. What are teachers' perceptions of TOEFL Junior Standard as a tool for student placement and monitoring student progress?
3. What are students' and teachers' perceptions of factors such as test difficulty, test length, instructional clarity, time limit, and test-taking experience?

Method

Context

The participants in this study were students and teachers at an afterschool EFL program at the Center for Research in Applied Linguistics of the American University of Armenia. The afterschool program aims to develop English language communication skills of school children in Armenia and has classes starting from A1 through B2 levels of proficiency based on CEFR. The duration of one term is 10 weeks: 20 hours of instruction for lower proficiency students and 40 hours for higher proficiency students. At the time of the study, the TOEFL Junior Standard had not been previously administered in Armenia, and the afterschool program students were not familiar with the test. The afterschool had been using a placement test from a textbook publisher for student placement purposes.

Participants

All students who registered for the Winter and Spring 2017 terms of the English language afterschool program took the TOEFL Junior Standard for placement at the time of registration. At the end of each test administration session, when all the test booklets and answer sheets had been collected, the students were asked to complete the student questionnaire. The usable response rate constituted 67% (202 of the total 322 students taking the TOEFL Junior Standard). The students were between ages 11 and 16,[1] with the average age of 13.7 years ($SD = 1.44$). There were more females than males (69% vs. 31%). The average placement score of the students was 744 ($SD = 76.5$) on the 600–900 total score scale of TOEFL Junior Standard (for details on placement distribution, see Appendix A).

In addition, all teachers of the afterschool program involved in teaching the above-mentioned subset of students were asked to participate in the study and complete the teacher questionnaire at the end of the Spring 2017 term. Nine teachers (eight female and one male), of whom two were also coordinators of the program, participated in the study. Given that the teachers were not familiar with the test, they had an initial introduction session where they looked at TOEFL Junior Standard test samples and familiarized themselves with the test-administration procedures as well as score and subscore descriptions. All the teachers acted as proctors during the two test-administration sessions.

Instruments

Two questionnaires (student and teacher) were administered to collect data for this study. The student questionnaire (adapted from Malone & Montee, 2014) had one open-ended and 11 closed (3- and 5-point Likert scale) items that aimed to investigate perceptions of (a) what TOEFL Junior Standard measures, (b) whether its difficulty and length seem appropriate, (c) whether instructions seem clear, (d) whether time seems sufficient, and (e) whether test performance is related to prior test-taking experience. The questionnaire was translated into Armenian to ensure that the wording of the questions was understandable to students age 11–16. Back-translation from Armenian into English was done to make sure that the underlying concepts of all items were retained in the translated version. The questionnaire was piloted with a sample of 20 respondents, after which some final revisions were made to the questionnaire items. Upon data collection, 11 surveys were discarded because they were blank. The student questionnaire measures, as presented in Appendix B, were as follows:

1. What TOEFL Junior Standard measures. This is a four-item scale (e.g., "The Listening section of the TOEFL Junior Standard allowed me to show how well I can listen in English") with *strongly agree* (5)*, agree, neither agree nor disagree, disagree, strongly disagree* (1)*,* and *do not know* options.
2. Time allotted for each section. This is a three-item scale (e.g., "I had enough time to answer the questions on the Listening section") with *strongly agree* (5)*, agree, neither agree nor disagree, disagree, strongly disagree* (1)*,* and *do not know* options.
3. Clarity of TOEFL Junior Standard instructions for students. This is a single-item measure ("I was able to easily understand the instruction given in the TOEFL Junior Standard") with *strongly agree* (5)*, agree, neither agree nor disagree, disagree, strongly disagree* (1)*,* and *do not know* options.
4. Length of TOEFL Junior Standard. This is a single-item measure ("What do you think about the length of the TOEFL Junior Standard?") with *it was too long* (3)*, it was just right,* and *it was too short* (1) options.

5. Difficulty of TOEFL Junior Standard. This is a single-item measure ("What do you think about the difficulty of the TOEFL Junior Standard?") with *it was too hard* (3), *it was just right,* and *it was too easy* (1) options.
6. Experience of taking TOEFL Junior Standard. This is a single-item measure of students' first exposure to TOEFL Junior Standard ("In general, my experience taking this test was") with *very positive* (5), *positive, neutral, negative,* and *very negative* (1) options and an open-ended item to explain the choice.

The teacher questionnaire (adapted from Malone & Montee, 2014) was administered at the end of the Spring 2017 term to collect data on teachers' perceptions of TOEFL Junior Standard as a placement and progress-monitoring tool. The questionnaire also included questions about what TOEFL Junior Standard measures, clarity of test instructions and test-administration instructions, clarity of the scores and subscores, and factors affecting student test performance. It had one open-ended and 17 closed (4- and 5-point Likert scale) items. A total of nine questionnaires were collected. The teacher questionnaire measures, as presented in Appendix C, were as follows:

1. What TOEFL Junior Standard measures. This is a four-item scale (e.g., "The Listening section of the TOEFL Junior Standard allows students to show how well they can listen in English") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
2. TOEFL Junior Standard for placement. This is a single-item measure ("The TOEFL Junior Standard is an accurate predictor of how well a student will perform in an EFL classroom") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
3. TOEFL Junior Standard for progress monitoring. This is a two-item measure (e.g., "The TOEFL Junior Standard is a useful tool for monitoring overall student progress in my classes") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
4. TOEFL Junior Standard score and subscore meaning. This is a two-item measure (e.g., "I understand what the TOEFL Junior Standard scores mean") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
5. Clarity of TOEFL Junior Standard instructions for students. This is a single-item measure ("TOEFL Junior Standard instructions for students are clear") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
6. Clarity of TOEFL Junior Standard administration instructions. This is a single-item measure ("The instructions for test administration are clear") with *strongly agree* (5), *agree, neither agree nor disagree, disagree, strongly disagree* (1), and *do not know* options.
7. Factors affecting students' performance on TOEFL Junior Standard. Six items were selected to measure the possible factors affecting students' performance of TOEFL Junior Standard based on teachers' own observations and, sometimes, students' reports. The teachers were asked to indicate the degree to which factors such as time pressure, test length, students' test anxiety, unfamiliarity of topics, distraction caused by other test takers, and difficulty of the language on the test could have affected performance. These factors had been identified during presurvey discussions with teachers. A 4-point Likert-type scale, ranking from *not at all* (1), *to a small extent, to some extent,* and *to a great extent* (4), was used.

The data analyses entailed computing frequency counts and means with standard deviations as well as content analysis of open-ended items.

Findings and Discussion

The present study investigated students' and teachers' perceptions of TOEFL Junior Standard. The results are presented by research question.

RQ 1: What Are Students' and Teachers' Perceptions of What the TOEFL Junior Standard Test Measures?

Using a 5-point scale from 1 (*strongly disagree)* to 5 (*strongly agree*), the results of the analysis of student' perceptions with regard to what TOEFL Junior Standard measures demonstrated that overall students agreed with the statements that the listening, reading, and language form and meaning sections of the TOEFL Junior Standard test allowed them to show how

Table 1 Students' Perceptions of What TOEFL Junior Standard (TJS) Measures (*N* = 202)

| | Frequency (percentage %) | | | | | | | |
Prompt	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree[a]	Do not know	*M*	*SD*
The Listening section of TJS allowed me to show how well I can listen in English.	53 (26%)	93 (46%)	36 (18%)	5 (3%)	4 (2%)	11 (5%)	3.97	0.87
The Reading section of TJS allowed me to show how well I can read in English.	50 (25%)	94 (47%)	38 (20%)	8 (4%)	2 (1%)	10 (5%)	3.95	0.85
The Language Form and Meaning section of TJS allowed me to show how well I can use English language vocabulary.	56 (28%)	99 (49%)	30 (15%)	5 (2%)	1 (0.5%)	11 (5%)	4.07	0.77
The Language Form and Meaning section of TJS allowed me to show how well I can use English grammar correctly.	45 (22%)	115 (57%)	29 (15%)	3 (1%)	2 (1%)	8 (4%)	4.02	0.73

[a]Strongly disagree = 1; strongly agree = 5.

Table 2 Teachers' Perceptions of What TOEFL Junior Standard (TJS) Measures (*N* = 9)

| | Frequency (percentage %) | | | | | | | |
Prompt	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree[a]	Do not know	*M*	*SD*
The Listening section of TJS allows students to show how well they can listen in English.	3 (33%)	6 (67%)	0 (0%)	0 (0%)	0 0%)	0 (0%)	4.33	0.50
The Reading section of TJS allows students to show how well they can read in English.	4 (44%)	5 (56%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4.44	0.53
The Language Form and Meaning section of TJS allows students to show how well they can use English vocabulary.	2 (22%)	5 (56%)	1 (11%)	0 (0%)	0 (0%)	1 (11%)	4.12	0.64
The Language Form and Meaning section of TJS allows students to show how well they can use English grammar correctly.	2 (22%)	5 (56%)	1 (11%)	0 (0%)	0 (0%)	1 (11%)	4.12	0.64

[a]Strongly disagree = 1; strongly agree = 5.

well they could listen and read in English as well as use English language vocabulary and grammar correctly (Table 1). All four questions received high average ratings (3.97, 3.95, 4.07, and 4.02, respectively), indicating that, on average, TOEFL Junior Standard was perceived by students to accurately represent their English language abilities.

These findings were further supported by the results of teachers' perceptions of the test (Table 2). The mean ratings of the four questions on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*), referring to the listening, reading, and language form and meaning sections of TOEFL Junior Standard (4.44, 4.33, 4.12, and 4.12, respectively), indicated that the teachers, on average, agreed with the statements that the sections of TOEFL Junior Standard were able to show how well their students could listen and read in English and use English language vocabulary and grammar correctly.

Table 3 Teachers' Perceptions of TOEFL Junior Standard (TJS) for Placement and Progress Monitoring (*N* = 9)

Prompt	Frequency (percentage %)					*M*	*SD*
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree[a]		
TJS is an accurate predictor of how well a student will perform in an EFL classroom.	1 (11%)	7 (78%)	0 (0%)	1 (11%)	0 (0%)	3.89	0.78
TJS is a useful tool for monitoring overall student progress in my classes.	3 (33%)	6 (67%)	0 (0%)	0 (0%)	0 (0%)	4.33	0.50
TJS helps me decide whether a student in my class is ready to go to the next level.	2 (22%)	2 (22%)	3 (44%)	1 (11%)	0 (0%)	3.56	1.01
I understand what the TJS scores mean.	6 (67%)	3 (33%)	0 (0%)	0 (0%)	0 (0%)	4.67	0.50
I understand what the TJS subscores mean.	4 (44%)	3 (33%)	1 (11%)	1 (11%)	0 (0%)	4.11	1.05

Note. EFL = English as a foreign language.
[a]Strongly disagree = 1; strongly agree = 5.

The results of the analysis of both student and teacher responses suggested that students' performance on TOEFL Junior Standard test tasks were perceived to be related to their performance in real-life classroom situations, with language measured by TOEFL Junior Standard being representative of the language actually used in class. As described in "*TOEFL Junior*® Design Framework" (So et al., 2017), during the development of TOEFL Junior Standard, each test task was designed to represent target language use tasks to be performed by young adolescent learners in English-medium secondary school contexts in order to ensure valid interpretations of test takers' language proficiency (Bachman & Palmer, 1996, 2010). The above-mentioned correspondence found between the test construct and teachers' and students' perceptions appeared to support the claim that test takers' performance on the TOEFL Junior Standard test "represents their relevant knowledge, skills, and abilities in their target domain—namely, their English use in classrooms and other learning contexts" (Butler, 2017, p. 261). This finding can provide additional evidence for the validity of TOEFL Junior Standard, as user beliefs about the match between test tasks and the skills they claim to measure provide substantial evidence for test validity (Malone & Montee, 2014).

RQ 2: What Are Teachers' Perceptions of TOEFL Junior Standard as a Tool for Student Placement and Monitoring Student Progress?

Using a 5-point scale from 1 (*strongly disagree)* to 5 (*strongly agree*), the analysis of teachers' responses to the questionnaire items related to the ability of TOEFL Junior Standard to place students and monitor student progress in English language classes suggested that teachers considered TOEFL Junior Standard an effective tool for predicting future student performance in an EFL classroom—for recording change in English language proficiency over time and helping them decide whether a student was ready to go to the next level (Table 3). Out of nine teachers, three teachers strongly agreed and six teachers agreed with TOEFL Junior Standard being able to document overall student progress in their classes, which suggested that TOEFL Junior Standard scores were perceived to capture improvement resulting from relevant learning experiences. This finding was further supported by teachers' ratings of the clarity of TOEFL Junior Standard score and subscore reports presented in Table 3 and their subsequent comments supplementing the provided ratings (e.g., Teacher A: "CEFR levels allowed me to understand what scores and subscores mean;" Teacher B: "Score descriptors helped me see my students' weaknesses;" Teacher C: "I showed to my students the descriptors and what skills they needed to develop;" and Teacher D: "My students were very excited to read about their strengths; I think this was motivating.. ."). These findings can serve as additional validity evidence, as they suggest that the test construct was clear to the teachers.

Table 4 Students' (*N* = 202) and Teachers' (*N* = 9) Perceptions of TOEFL Junior Standard (TJS) Instructions

Group	Prompt	Frequency (percentage %)						*M*	*SD*
		Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree[a]	Do not know		
Students	I was able to easily understand TJS instructions.	90 (47%)	78 (39%)	21 (10%)	3 (1%)	1 (0.5%)	9(4%)	4.31	0.77
Teachers	TJS instructions for students are clear.	6 (67%)	3 (33%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4.67	0.50
	TJS instructions for test administration are clear.	8 (89%)	1 (11%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4.78	0.67

[a]Strongly disagree =1; strongly agree = 5.

Table 5 Students' Perceptions of Time Allotted for Each Section (*N* = 202)

Prompt	Frequency (percentage %)						*M*	*SD*
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree[a]	Do not know		
I had enough time to answer the questions on the Listening section.	37 (19%)	83 (41%)	55 (27%)	21 (10%)	4 (2%)	2 (1%)	3.64	0.97
I had enough time to answer the questions on the Reading section.	56 (28%)	89 (44%)	40 (20%)	14 (7%)	2 (1%)	1 (0.5%)	3.91	0.92
I had enough time to answer the questions on the Language Form and Meaning section.	63 (31%)	83 (41%)	38 (19%)	14 (7%)	2 (1%)	2 (1%)	3.96	0.94

Note. Strongly disagree = 1; strongly agree = 5.

RQ 3: What Are Students' and Teachers' Perceptions of Factors Such as Test Difficulty, Test Length, Instructional Clarity, Time Limit, and Test-Taking Experience?

Further, taking into consideration the potential negative impact of the language of test instructions and time pressure on students' test performance, it was necessary to understand students' and teachers' perceptions of the level of clarity of test instructions and time pressure with regard to TOEFL Junior Standard. The average ratings on a 5-point Likert scale, from 1 (*strongly disagree)* to 5 (*strongly agree*), presented in Table 4 demonstrate that both students and teachers judged the test instructions to be clear and easy to understand (4.31 and 4.67, respectively). The teachers were also asked to rate the instructions for test administration, and eight out of nine teachers strongly agreed with the instructions being clear.

In addition, as Table 5 demonstrates, students in general believed that the time allotted for completing the sections of TOEFL Junior Standard was enough. Although the listening section was the one that received the lowest average rating in all age groups, only 25 out of 200 students rated the time for completing the listening section as not enough.

The frequencies and percent of students' perceptions of the length and difficulty of TOEFL Junior Standard, presented in Table 6, suggested that the majority of the students thought the test had the right length (83%) and level of difficulty (90%). These findings suggested that the content and format of TOEFL Junior Standard test tasks were developmentally appropriate for the students, allowing them to demonstrate their target abilities (Hasselgreen, 2005; McKay, 2006).

As demonstrated by teachers' ratings on a 4-point scale, ranking from 1 (*not at all*) to 4 (*to a great extent*), the length of TOEFL Junior Standard was reported to have the greatest negative impact on students' test performance (Table 7). According to teachers' comments, filling in the background information on TOEFL Junior Standard answer sheets took too long (e.g., Teacher E: "The only problem that students mentioned a lot was the time spent on filling in the paper about their personal information;" Teacher A: "Even before the actual test started they were already very tired;" and Teacher A and Teacher D: "It took very long to fill in the personal information; everyone got tired.").

Table 6 Frequency (and Total Percent) of Students' Responses on Perceived Length and Difficulty of TOEFL Junior Standard ($N = 196$)

	Test length			Test difficulty		
Age	Too short $n = 6$ (3%)	Just right $n = 163$ (83%)	Too long[a] $n = 27$ (14%)	Too easy[b] $n = 8$ (4%)	Just right $n = 177$ (90%)	Too difficult $n = 11$ (6%)
11	0	14	2	0	16	0
12	1	20	9	0	25	5
13	0	33	5	3	32	3
14	0	38	7	1	43	1
15	3	44	4	4	45	2
16	2	14	0	0	16	0

[a]Too short $= 1$; too long $= 3$. [b]Too easy $= 1$; too difficult $= 3$.

Table 7 Teachers' Perceptions of Factors Affecting Students' Performance on TOEFL Junior Standard ($N = 9$)

	Frequency (percentage %)					
Factor	Not at all	To a small extent	To some extent	To a great extent[a]	*M*	*SD*
Time pressure	0 (0%)	1 (11%)	5 (56%)	3 (33%)	3.22	0.67
Test length	1 (11%)	0 (0%)	2 (22%)	6 (67%)	3.44	1.01
Students' test anxiety	1 (11%)	0 (0%)	4 (44%)	4 (44%)	3.22	0.97
Unfamiliarity of topics	2 (22%)	1 (11%)	3 (33%)	3 (33%)	2.78	1.20
Distraction caused by other test takers	2 (22%)	5 (56%)	0 (0%)	2 (22%)	2.22	1.09
Difficulty of the language on the test	1 (11%)	5 (56%)	2 (22%)	1 (11%)	2.33	0.87

[a]Not at all $= 1$; to a great extent $= 4$.

Table 8 Students' Perceptions of Test-Taking Experience

	Test-taking experience						
Age	Negative $n = 1$ (0.5%)	Neutral $n = 23$ (12%)	Positive $n = 106$ (54%)	Very positive[a] $n = 65$ (33%)	Total	*M*	*SD*
11	0	4	10	1	15	3.80	0.56
12	1	5	13	11	30	4.13	0.82
13	0	4	22	12	38	4.21	0.62
14	0	4	30	11	45	4.16	0.56
15	0	5	24	22	51	4.33	0.65
16	0	1	7	8	16	4.44	0.63

[a]Very negative $= 1$; very positive $= 5$.

Using a 5-point scale, from 1 (*very negative*) to 5 (*very positive*), students rated their experience of taking the TOEFL Junior Standard test and explained their ratings (open-ended item) in the last two questions. Table 8 shows that the test-taking experience was almost uniformly rated as positive. In addition, a total of 183 students responded to the open-ended item to explain their choice for the previous question. Responses that were illegible or contained only the word *positive* were deleted, leaving 168 responses. Each response was examined to determine whether it explained the student's choice for the previous question and was coded based on the categories developed after the preliminary review. A total of 19 students (11%) mentioned solely the word *interesting*. The remaining 149 responses were coded by two coders with a good intercoder agreement as determined by Cohen's kappa, $\kappa = .792$, 95% CI [.669, .915], $p < .001$. Thus, according to the analysis, this positive experience was related to the opportunity to test their knowledge of English (64%) and the opportunity to gain test-taking experience (36%).

The positive ratings of experience of taking TOEFL Junior Standard suggested that the test tasks were engaging, which is essential in young learners' assessment, taking into consideration the importance of affective factors such as motivation

and engagement in young learner test performance (Wolf & Butler, 2017). The teachers (five out of nine) in their comments also noted that students, and especially their parents, particularly valued the fact that "TJS is an internationally recognized test," provides "an independent evaluation of English language proficiency," "can be submitted as an official score," and "can serve as official evidence of English language level."

Overall, the findings of the study suggested that both students and teachers perceived TOEFL Junior Standard as an effective measure of English language proficiency. As such, these findings are in line with the results of validation studies on TOEFL Junior Standard (e.g., Gu et al., 2015; Papageorgiou & Cho, 2014; Wolf & Steinberg, 2011) that support the use of TOEFL Junior Standard for placement decisions and for monitoring English language development in young language learners over time. The findings of the current study offer stakeholder perspectives on the TOEFL Junior Standard test critical in establishing a link between test validation and test use (Cheng & DeLuca, 2011; Moss et al., 2006) and thus can have implications for test developers and other relevant stakeholders.

In addition, most research on TOEFL Junior Standard has been conducted in ESL settings. Considering that the number of *TOEFL iBT*® takers in EFL settings by far exceeds those in ESL settings (ETS, 2018), it is critical to conduct additional research on TOEFL tests in EFL settings to make the findings more representative internationally. Because this study took place in a developing and non-English speaking country, it adds a new understanding of how young learners of English and their teachers in an EFL setting view TOEFL Junior Standard. Comparisons of TOEFL Junior Standard stakeholder perceptions across EFL settings and between EFL and ESL contribute to the test validity evidence because English learning opportunities and assessment formats often differ. For example, Cho and So (2014) showed that their Korean young learners of English identified more challenges with TOEFL Junior Standard than the Armenian participants in this study. This difference may be due to the fact that the Korean study was based on in-depth interviews with a small sample of participants versus a survey methodology in the current study. However, similar studies, especially in EFL contexts, number too few to identify any meaningful patterns. Hence, more research on TOEFL Junior Standard stakeholder perceptions in EFL settings is important.

Limitations and Directions for Future Research

The findings of the present study must be interpreted in light of its limitations. First, the sample of students from a particular afterschool program limits the generalizability of the findings to other teaching and learning contexts. Future research should include a more diverse sample of EFL students and also a larger and more diverse sample of EFL teachers. Subsequent research should also examine test takers' perceptions of the paper-based (standard) versus the computer-delivered (comprehensive) modes of TOEFL Junior, focusing on test takers' attitudes to the test delivery mode. This information would provide a more comprehensive understanding of which delivery mode is more desirable for today's young EFL learners.

Conclusion

This study adds to the growing research on the validation and use of the TOEFL Junior Standard test. The findings contribute to the understanding of stakeholder perceptions of TOEFL Junior Standard, suggesting that the perceptions of TOEFL Junior Standard in general are positive, and the test is well received, both by the student test takers and their teachers. The results demonstrate that both students and teachers in general agree that TOEFL Junior Standard tasks are effective in measuring students' language proficiency. The TOEFL Junior Standard test tasks are perceived to be developmentally appropriate for the students, allowing them to demonstrate their English language abilities. These findings can be viewed as additional evidence for the validity of the TOEFL Junior Standard test, as they are indicative of the correspondence between the TOEFL Junior Standard test construct and test users' perceptions and thus support the use of TOEFL Junior Standard for student placement and monitoring growth in students' English language proficiency in EFL settings.

Note

1 The top age range is 16 here because some of the 15-year-old students turned 16 during the study.

References

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.

Aydin, S. (2012). The effects of young EFL learners' perceptions of tests on test anxiety. *Education3–13,* 40, 189–204. https://doi.org/10.1080/03004279.2010.510119

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*, 1–34. https://doi.org/10.1207/s15434311laq0201_1

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford, England: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford, England: Oxford University Press.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Longman.

Butler, Y. (2017). Challenges and future directions for young learners' English language assessments and validity research. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 254–274). New York, NY: Routledge. https://doi.org/10.4324/9781315674391

Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment, 16*, 104–122. https://doi.org/10.1080/10627197.2011.584042

Chik, A., & Besser, S. (2011). International language test taking among young learners: A Hong Kong case study. *Language Assessment Quarterly, 8*, 73–91. https://doi.org/10.1080/15434303.2010.537417

Cho, Y., & So, Y. (2014). *Construct-irrelevant factors influencing young English as a foreign language (EFL) learners' perceptions of test task difficulty* (Research Memorandum No. RM-14-04). Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2018). *Test and score data summary for* TOEFL iBT® *tests*. Retrieved from https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf

Green, A. (2012). Placement testing. In C. Coombe, B. O'Sullivan, P. Davidson, & S. Stoynoff (Eds.), *The Cambridge guide to language assessment* (pp. 164–170). Cambridge, England: Cambridge University Press.

Gu, L., Lockwood, J., & Powers, D. E. (2015). *Evaluating the* TOEFL Junior® *Standard test as a measure of progress for young English language learners* (Research Report No. RR-15-22). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12064.

Gu, L., & So, Y. (2015). Voices from stakeholders: What makes an academic English test 'international'? *Journal of English for Academic Purposes, 18*, 9–24. https://doi.org/10.1016/j.jeap.2014.10.002

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing 22*(3)*,* 337–354.

Hawkey, R. A. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge, England: Cambridge University Press.

Malone, M. E., & Montee, M. (2014). *Stakeholders' beliefs about the* TOEFL iBT® *test as a measure of academic language ability* (Research Report No. RR-14-42). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12039

McKay, P. (2006). *Assessing young language learners*. Cambridge, England: Cambridge University Press.

McNamara, T. (2007). Language testing: A question of context. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 133–140). Ottawa, Canada: University of Ottawa. https://doi.org/10.2307/j.ctt1ckpccf.13

Messick, S. (1990). *Validity of test interpretation and use* (Research Report No. RR-90-11). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1990.tb01343.x.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30,* 109–162. https://doi.org/10.3102/0091732X030001109

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*, 223–239. https://doi.org/10.1177/0265532213499750

Saville, N. D. (2003) The process of test development and revision within UCLES EFL. In C. J. Weir & M. Milanovic (Eds), *Continuity and innovation: Revising the Cambridge proficiency in English examination 1913–2002: Vol. 15. Studies in language testing* (pp. 57–120). Cambridge, England: Cambridge University Press.

Saville, N. D. (2010). Developing a model for investigating the impact of language assessment. *Research Notes, 42*, 2–8.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*, 298–317. https://doi.org/10.1177/026553229601300305

So, Y., Wolf, M. K., Hauck, M. C., Mollaun, P., Rybinski, P., Tumposky, D., & Wang, L. (2017). *TOEFL Junior*® design framework. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 59–78). New York, NY: Routledge. https://doi.org/10.4324/9781315674391

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499–509. https://doi.org/10.1016/S0346-251X(00)00035-X

Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly, 45*, 628–660. https://doi.org/10.5054/tq.2011.268063

Wolf, M. K., & Butler, Y. G. (2017). An overview of English language proficiency assessments for young learners. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 1–21). New York, NY: Routledge. https://doi.org/10.4324/9781315674391

Wolf, M. K., & Steinberg, J. (2011). *An examination of United States middle school students' performance on* TOEFL Junior® (Research Memorandum No. RM-11-15). Princeton, NJ: Educational Testing Service.

Appendix A

Students' Placement Distribution

CEFR level	Age						Total
	11	12	13	14	15	16	
A1	4	5	3	3	4	1	20
A2	7	18	9	7	7	3	51
B1-	6	7	18	20	17	4	72
B1	0	1	6	12	15	4	38
B2	1	0	2	5	8	5	21
Total	18	31	38	47	51	17	202

Appendix B

Student Questionnaire

1. Please indicate your agreement with the following statements.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Do not know
The Listening section of the TOEFL Junior Standard let me show how well I can listen in English.						
The Reading section of the TOEFL Junior Standard let me show how well I can read in English.						
The Language Form and Meaning section of the TOEFL Junior Standard let me show how well I can use English vocabulary.						
The Language Form and Meaning section of the TOEFL Junior Standard let me show how well I can use English grammar correctly.						
I was able to easily understand the instruction given in the TOEFL Junior Standard.						
I had enough time to answer the questions on the Listening section.						
I had enough time to answer the questions on the Reading section.						
I had enough time to answer the questions on the Language Form and Meaning section.						

2. What do you think about the length of the TOEFL Junior Standard?

    a. It was too long
    b. It was just right
    c. It was too short

3. What do you think about the difficulty of the TOEFL Junior Standard?

    a. It was too hard
    b. It was just right
    c. It was too easy

4. In general, my experience taking this test was

    a. Very positive
    b. Positive
    c. Neutral
    d. Negative
    e. Very negative.

5. Please explain in one sentence your answer to Question 4 above:

    _____

    _____

Appendix C

Teacher Questionnaire

1. Please indicate your agreement with the following statements

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Don't know
The Listening section of the TOEFL Junior Standard allows students to show how well they can listen in English.						
The Reading section of the TOEFL Junior Standard allows students to show how well they can read in English.						
The Language form and Meaning section of the TOEFL Junior Standard allows students to show how well they can use English vocabulary.						
The Language Form and Meaning section of the TOEFL Junior Standard allows students to show how well they can use English grammar correctly.						
The TOEFL Junior Standard is an accurate predictor of how well a student will perform in an EFL classroom.						
The TOEFL Junior Standard is a useful tool for monitoring overall student progress in my classes.						
The TOEFL Junior Standard helps me decide whether a student in my class is ready to go to the next level.						
I understand what the TOEFL Junior Standard scores mean.						
I understand what the TOEFL Junior Standard subscores mean.						
The test instructions for students are clear.						
The instructions for test administration are clear.						

2. Based on your observations and/or student reports, to what extent do any of the following factors affect students' performance on the TOEFL Junior Standard?

	To a great extent	To some extent	To a small extent	Not at all	Don't know
Time pressure					
Test length					
Students' test anxiety					
Unfamiliarity of topics					
Distraction caused by other test takers					
Difficulty of the language on the test					

3. Please use the space below to comment on your responses:

_____

_____

_____

_____

_____

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/