

Test Taker Attitudes to Response Time Length in Speaking Tests

Kristen Sullivan

Shimonoseki City University

Reference Data:

Sullivan, K. (2013). Test taker attitudes to response time length in speaking tests. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. Tokyo: JALT.

This paper forms part of a larger study concerned with the effect of response time length on responses to TOEFL iBT independent speaking tasks. Test takers are currently given 45 seconds to complete their responses to independent speaking questions. However, given the nature of spontaneous interactive speech, I question whether 45 seconds is indeed enough time for test takers to fully develop their response and demonstrate their best ability. In this study, 36 university students responded to 3 independent speaking test tasks that were allocated different response time lengths (45 seconds, 90 seconds, and 135 seconds). Participants also completed 2 sets of surveys designed to question their attitudes toward the tasks, their performance on the tasks, and their preferences regarding response time length. In this paper I look specifically at the survey results and report on test taker attitudes to TOEFL iBT independent speaking task response time lengths.

本論文は、TOEFL iBTのIndependent Speaking (IS)問題において解答時間が解答にどのような影響があるかを調べる研究調査の一部を成している。現在、IS問題の解答時間が45秒で設定されている。しかし、即興発話の性質を考慮すると、受験者が実力を出すのに45秒が果たして十分だと言えるかどうか。よって、本論文はTOEFLのIS問題の解答時間の妥当性を調査対象とする。本調査では、36人の大学生が3つのIS問題に対して解答した。それぞれの問題に異なる解答時間（45秒、90秒、135秒）が割り当てられた。また、学生は、テスト問題に対する考え方や解答の出来具合に対する反応、解答時間に対する希望を尋ねる2種類のアンケートにも答えた。本論文はこの2種類のアンケートの結果を報告し、受験者のIS問題の解答時間に対する考え方を調べる。

THE PURPOSE of this paper is to report on the initial findings of a study that is investigating the influence of response time length on responses made to TOEFL iBT (Test of English as a Foreign Language Internet-Based Test) independent speaking tasks. The larger study is investigating the effect of three different response times on test taker scores, the discourse of the responses, and test taker attitudes toward the tasks and task conditions, with the aim to identify if and how response time length influences performance. This paper includes an initial descriptive analysis of the results of two sets of surveys that investigated test taker responses to their performance on speaking tasks taken under different response time conditions, and their overall attitudes toward response time length for speaking tests.



The TOEFL iBT Test

The TOEFL iBT test is a high stakes, gate-keeping test of academic English proficiency. For students at Japanese universities, TOEFL iBT scores are commonly used as a language requirement benchmark for participation in study abroad programs at English-language universities. The Liberal Democratic Party also recently proposed that the TOEFL test be used for admission to public universities (Yoshida, 2013). The TOEFL iBT test is a highly researched exam (c.f., Chapelle, Enright, & Jamieson, 2008) that was introduced in 2005 as the successor of the TOEFL CBT (Computer-Based Test) and TOEFL PBT (Paper-Based Test). As its name suggests, the TOEFL iBT test is conducted completely by computer, with test items and responses transferred via the Internet for administration and scoring. The test is made up of four sections (reading, listening, speaking, and writing) and the speaking section, which is the focus of this paper, consists of two main types of test items: independent tasks (which require the test taker to respond to a prompt that asks them to give and explain their opinion) and integrated tasks (which require test takers to summarize written and spoken texts).

The TOEFL iBT test is the first version of the TOEFL test to include a speaking section. In the speaking section, test instructions and questions are presented to the test taker via the test station computer (both on screen and via audio recordings) and test-taker spoken responses are automatically recorded by the test station. The TOEFL iBT speaking section can thus be categorized as a semi-direct test of spoken English.

TOEFL iBT Speaking Tasks: A Background to Response Time Lengths

In this study I focused on the independent speaking tasks, in particular the free-choice speaking tasks. In these tasks, test takers are given 15 seconds to prepare and 45 seconds to give their

response to a prompt that appears on their test station computer screen. The preparation and response times are shown to test takers via a countdown clock that also appears on their computer screen.

The concern I wanted to address is whether the 45-second response time limit is enough to allow test takers to demonstrate their best ability, especially in consideration of the nature of spontaneous interactive speech which is characterized by hesitations, self-repair, and “constraints of breath and spoken language processing” (Hughes, 2002, p. 77). The issue of response time length is important if we are concerned with creating test tasks which “bias for best” possible performance (Swain, 1984, pp. 195-196). It must also be considered from the viewpoint of achieving task authenticity and maintaining a positive impact on how teachers and learners practice speaking in the classroom. It is also a potential factor that can influence test taker perceptions of task difficulty and authenticity—that is, their perceptions of the face validity of the test—which can affect how the test taker responds to test items and whether they accept their test results (c.f., Wigglesworth & Elder, 2010).

While response time length does appear to be an important test condition for speaking tests, a review of the literature reveals that there is a dearth of research on the topic. Rather, the effect of planning time length has been the predominate focus of research to date. This can most probably be explained by the fact that interview tests (i.e., direct speaking tests) are more commonly in use for speaking tests rather than semi-direct tests such as the TOEFL iBT test.

While not a main focus of research, the issue of the response time length of TOEFL speaking tasks was investigated by the ETS (Educational Testing Service) in the development stages of the TOEFL iBT test. An early theoretical study concluded that the collection of larger samples of the test taker’s oral proficiency would be desirable and that test takers “should have the op-

portunity to produce a total of at least 10 to 15 minutes' worth of speech for assessment" (Butler, Eignor, Jones, McNamara, & Suomi, 2000, p. 13). How the 10 to 15 minutes should be divided and allocated among the various test items in the speaking section, however, was not touched upon.

In a prototype study looking at integrated speaking test items for the TOEFL iBT test, Enright, Bridgeman, Eignor, Lee, and Powers (2008) sought to investigate whether test scores were affected by different response lengths. They found that while mean test scores were slightly higher for test takers who had 120 seconds (versus 60 seconds) and 90 seconds (versus 150 seconds) of response time, this difference was not statistically significant. This finding was used to justify the prioritizing of factors such as "domain coverage, expert opinion, availability of text materials, and cost of development" over issues of "task characteristics and administration conditions" in iBT speaking test development (Enright et al., 2008, p. 128). Indeed, the shortest response time length of 60 seconds is now in use for the integrated speaking tasks in the TOEFL iBT test.

These kinds of decisions are important for test creators who also need to consider the human and monetary resources involved in the implementation of their test items (c.f., Bachman and Palmer, 1996). The testing of speaking ability is indeed resource intensive. However, this balancing act needs to be conducted in a way that does not negatively affect the test taker's performance on and experience of the test. In regard to the TOEFL iBT speaking section this is an issue that deserves deeper consideration.

Study Aims and Design

To further investigate these issues, I aimed in this study to replicate the Enright et al. (2008) study by conducting a similar experiment, this time looking at the effect of response time

length on responses to free-choice independent speaking test items. The lack of past research gave few clues to guide the choice of response time lengths to be used in this study. However, it made sense to compare the current time of 45 seconds with response times of double (90 seconds) and triple (135 seconds) this amount of time. In addition, coming from an understanding that it is important to also consider test takers' perceptions of the test and test conditions in test development and validation endeavors (Wigglesworth & Elder, 2010) participants also responded to a series of surveys questioning their attitudes toward the speaking tasks.

The participants in this study were 36 undergraduate and study abroad students studying within the economics faculty of a small, public university in rural Japan. The participants were recruited through the use of on-campus advertisements and announcements, and participation was voluntary. The difficulty in finding participants within a nonlanguage department meant that all respondents regardless of year level, proficiency level, and past experience with the TOEFL test were invited to participate in the study. There were 16 male and 20 female participants. By year level, four were 1st-year students, five were 2nd-year students, 14 were 3rd-year students, nine were 4th-year students, and four were exchange students. There were 29 native Japanese speakers, one bilingual Japanese-Spanish speaker, five native Chinese speakers, and one Turkish speaker.

Participants were required to respond to three speaking tasks taken under three different response time conditions (45 seconds, 90 seconds, and 135 seconds). Response time length, speaking task questions, and the order in which these were presented to participants were organized into different combinations to take into account any effects of item topic and task order. The experiment conditions were designed to replicate those of the actual TOEFL iBT test as much as possible, and the test item prompts were taken from official TOEFL iBT practice

tests (Educational Testing Service, 2007; Educational Testing Service, 2009).

After each speaking task participants were asked to respond to a survey questioning their attitudes toward the task and their performance on the task. Furthermore, after completion of all three tasks participants responded to another survey about response time preferences. This paper offers a descriptive analysis of responses to these surveys to gain an initial understanding of test taker beliefs about response time lengths for speaking tasks. The paper will firstly introduce the findings of the most imperative questions asked in both surveys, before discussing these findings and their implications for the speaking section of the TOEFL iBT test and response time lengths in speaking tests in general.

Results

Immediate Posttask Survey Results

The following questions were asked in each immediate post-task survey. This section will descriptively analyze test-taker responses to these questions, comparing their responses by response time length.

1. Did you have enough time to answer the question?
2. Could you complete your response?
3. Was it difficult to respond to the question?
4. Were you satisfied with your response?
5. Could you demonstrate your ability?
6. If the response time were longer would you be able to give a better response?

Adequacy of Response Time Length to Answer Questions and Complete Responses

Being able to complete one's response is obviously important not just in regard to the test taker's eventual score, but also for how the test taker feels about the test item itself. As Table 1 shows, more than half of the participants felt that they had enough time to answer the question under each response time condition. However, the 90-second response time seems to have been most adequate, with 72% of respondents saying they had enough time under this condition. In contrast, 42% of students said they felt they did not have enough time under the 45-second condition, and 31% of students responded that they had too much time under the 135-second condition.

Table 1. Question 1 Results (Did you have enough time to answer the question?)

Response	45 seconds	90 seconds	135 seconds
Too much time	1 (3%)	5 (14%)	11 (31%)
Enough time	20 (56%)	26 (72%)	23 (64%)
Not enough time	15 (42%)	5 (14%)	2 (6%)

In response to question 2 that asked test takers if they could complete their response, slightly more than half of participants felt they could complete their response under the longer response time conditions. In contrast, slightly more than half of test takers felt they could not complete their response under the shorter 45-second condition (see Table 2).

Table 2. Question 2 Results (Could you complete your response?)

Response	45 seconds	90 seconds	135 seconds
Yes	15 (42%)	21 (58%)	19 (53%)
No	21 (58%)	15 (42%)	17 (47%)

Perceptions of Item Difficulty

In regard to test taker perceptions of item difficulty, as Table 3 shows, slightly more participants felt that the items with longer response time conditions were difficult to respond to. However, it is interesting to note that more than 50% of respondents felt that it was difficult to respond to each question, regardless of response time length.

Table 3. Question 3 Results (Was it difficult to respond to the question?)

Response	45 seconds	90 seconds	135 seconds
Yes	20 (56%)	22 (61%)	24 (67%)
No	16 (44%)	14 (39%)	12 (33%)

Satisfaction with Response and Performance

As Table 4 shows, the majority of students were dissatisfied with their response regardless of response time length. However, it is noteworthy that 83% of students were dissatisfied with their response made under the 90-second condition. This is interesting given that students were more likely to respond that they had enough time to answer the question and complete

their response under this time condition, which suggests that satisfaction with response may not be related to perceptions of having sufficient time to answer the test question.

Table 4. Question 4 Results (Were you satisfied with your response?)

Response	45 seconds	90 seconds	135 seconds
Yes	13 (36%)	6 (17%)	9 (25%)
No	23 (64%)	30 (83%)	27 (75%)

For this question participants were also asked to explain why they were satisfied or dissatisfied with their response. Participant comments were coded by the researcher and the results are summarized below. For each response time condition, satisfaction with the content of the response (i.e., being able to give enough reasons, being able to give a complete response, and being able to say what they wanted to say) was given as a reason. Only for the 45-second condition was using the response time efficiently (i.e., using up all of the response time and completing the response within the set response time) offered as a reason.

As for the reasons why participants were dissatisfied with their responses, for all response time conditions not being able to complete the response and not being able to give enough information were offered. Only for the 45-second condition was not being able to respond at all (i.e., stopping halfway and having no ideas to talk about) given as a reason, and only for the 90-second condition was having time left over or finishing the response too quickly mentioned.

Here it is interesting to see that test takers nominated response content rather than grammar, vocabulary, or pronunciation as a key reason for their satisfaction or dissatisfaction with

their response. The connection between task completion and satisfaction is also of interest. It is also noteworthy that students were concerned about how they used their response time, and that there was a perception that finishing early is not good. This is all important because reasons for test taker satisfaction or dissatisfaction are presumably connected to how test takers perceive their response will be scored, and these perceptions can affect how they approach making their responses, and in larger terms, how they actually prepare for the test.

For question 5, which asked test takers if they felt they could demonstrate their ability through the task, the response patterns were quite consistent over all response time conditions. Regardless of response time length, approximately two-thirds of test takers felt that they could not demonstrate their ability through the test task (see Table 5).

Table 5. Question 5 Results (Could you demonstrate your ability?)

Response	45 seconds	90 seconds	135 seconds
Yes	13 (36%)	14 (39%)	14 (39%)
No	23 (64%)	22 (61%)	22 (61%)

Potential to Improve Performance with Longer Response Time

Contrary to the very similar response patterns that we saw for question 5, for question 6, which asked participants if they could improve their performance with more response time, we see a split in responses between the 45-second condition and the 90-second and 135-second conditions. As Table 6 shows, after completing tasks under the longer response time lengths,

approximately two-thirds of test takers felt that they would not be able to do better with more time. In contrast, after making their response under the 45-second condition approximately two-thirds of test takers felt that they could improve with more time. This suggests that while test takers felt that 45 seconds was not enough time, they thought that a response time of 90 to 135 seconds was sufficient.

Table 6. Question 6 Results (If the response time were longer would you be able to give a better response?)

Response	45 seconds	90 seconds	135 seconds
Yes	23 (64%)	11 (31%)	12 (33%)
No	13 (36%)	23 (64%)	22 (61%)
Not sure	0 (0%)	2 (6%)	2 (6%)

For this question too, participants were asked to explain why they believed their response would or would not improve with more time. A common reason given for all time conditions was that they could not think of any further ideas to talk about. For the 90- and 135-second conditions a limitation in English ability was given as a reason. In addition, several participants responded that if anything they would prefer more preparation time.

For test takers who answered that their response would improve with more time, for the responses made under the 45- and 90-second conditions, being able to add more information was given as the primary reason, which parallels the importance given to response content raised in question 4 about response satisfaction. Interestingly, for both the shortest condition (45 seconds) and the longest condition (135 seconds) test takers

said they could improve their response with more time because they would be able to prepare or think more about their answer. Only for responses made under the 45-second condition was being more relaxed given as a reason, which suggests that for certain test takers the shorter time condition caused some level of anxiety.

Overall Attitudes to Response Time Lengths

After completion of all three tasks, students were asked to respond to a final questionnaire that surveyed their overall attitudes to the speaking tasks, particularly the response lengths. Here student responses to three pertinent questions asked in the survey will be examined:

1. Under which response time could you best demonstrate your ability?
2. For this kind of speaking test, how much response time do you need to best demonstrate your ability?
3. What do you think about having response time limits in speaking tests like this one?

Best Response Time Condition to Demonstrate Ability

Data in Table 5 shows that regardless of response time condition, participants were generally pessimistic about their performance with approximately two-thirds of test takers responding that they could not demonstrate their ability through the tasks. Thus, it is interesting that after completing all tasks the majority of test takers felt more positively about their performance under the longer time conditions with 56% responding that they could best demonstrate their ability under the 90-second condition and 28% under the 135-second condition (see Table 7).

Table 7. Response Time Length That Allowed Test Taker to Best Demonstrate Ability

45 seconds	90 seconds	135 seconds
6 (17%)	20 (56%)	10 (28%)

Test takers were also asked to explain their response to this question. For the 17% of students who nominated the 45-second time condition, the reason given was that they did not have that much to say anyway and would thus find it difficult to speak for longer. For the 90-second condition, feeling that they had enough time to complete their response, give reasons, and say what they wanted to say, without having too much time left over, were common reasons given by test takers. In addition to this, and in common with responses given by students who nominated the 135-second condition, were the ideas of having enough time to think about their response and not feeling pressured. This repeats the connection between response time length and anxiety levels first raised in question 6 in the immediate posttask survey. A common theme among these responses is the importance of having the appropriate amount of time to complete one's response. It seems that for students with not much to say this meant a shorter response time, and for students with more to say this meant a longer response time.

Preferred Response Time Length

This question asked test takers to nominate their own preferred response time length and to explain this preference. Test taker responses are summarized in Table 8.

Table 8. Preferred Response Time Length

Less than 60 seconds	60 seconds	90 seconds	Around 120 to 150 seconds
1 (4%)	7 (26%)	8 (30%)	11 (41%)

Note. For purposes of simplification, six responses are not included in Table 8: one participant responded 45 or 90 seconds; two participants responded 60 to 90 seconds; one participant responded 45 or 135 seconds depending on the topic; one participant responded more than 90 seconds; one participant responded 5 minutes. In addition, three participants did not respond with a particular time length; two of these participants responded that they would prefer more pretask preparation time. Thus for this question $n = 27$.

As Table 8 shows, 96% of test takers felt that 60 seconds or more is required to best demonstrate their ability under this type of task. The majority of test takers thus did not believe that the current response time length of 45 seconds was enough to do their best. The main reason given for this preference was that the longer response time lengths would enable online planning. Other reasons were that the longer response times would allow them not to feel pressured, to give more detail, to give a logical response, and to deal with problems with delivery.

Test Taker Opinions About Time Limits in Speaking Tests

Given the trend toward test taker preferences for longer response time lengths, one may assume that test takers would prefer speaking test tasks with no specific time limits. However, perhaps surprisingly, the majority of test takers responded that they do not mind the presence of time limits (see Table 9). Indeed, many replied that having a time limit gives an indication of examiner or scorer expectations, and thus provides them with a goal for the task. Another reason was that a time limit forces

the test taker to get to the point, again having a goal-setting function. Interestingly, several students also referred to issues of test practicality and fairness in their responses, which shows that test takers are very sophisticated in their understanding of the behind-the-scenes issues that go along with high stakes assessment.

Table 9. Test Taker Opinions about Time Limits in Speaking Tests

Time limits are okay	I would prefer no time limit	I don't care
25 (69%)	10 (28%)	1 (3%)

For the 28% of test takers who responded that they would prefer not to have a response time limit, the most common reason was that it is a distraction and puts pressure on the test taker, which is an issue that has already been touched upon above. Another reason was that the necessary amount of response time depends on the question itself, suggesting that test taker reactions to item content are also a factor that should be considered.

Discussion

These initial descriptive results suggest that most test takers are not satisfied with the current TOEFL iBT independent speaking response time limit of 45 seconds and believe that they can perform more strongly under longer response time conditions. It seems that for most test takers 45 seconds is not enough to make a complete response, especially in terms of depth of content. It also seems that for some, the shorter time length can result in enhanced feelings of anxiety, which could affect response qual-

ity in various ways. Concern about finishing early and having time left over is also an indication that many test takers believe that doing so will have a negative effect on their score; indeed, many noted that having a time limit gives an important indication of the type of response that is being asked of them. This is an important reminder of the need for test developers to clearly state expectations and criteria. If they do not, students will be forced to come up with their own conclusions, which may negatively impact how they go about responding to the task and actually preparing for the test.

Future Research

The results discussed in this paper are purely descriptive and must still be subjected to further statistical analysis. Moreover, it is important to remember that these results indicate only how students reacted to the item characteristics; analysis of the effect of response time on scores and actual performance is still to be conducted and will be reported in a future paper.

In their study on integrated speaking tasks, Enright et al. (2008) concluded that a lack of a statistically significant difference in scores for responses made under different time conditions allows test creators to prioritize practicality concerns when deciding test conditions. However, given the high stakes nature of the TOEFL iBT test and the fact that test taker attitudes toward a test may affect not only how they perform on the test day but also how they prepare for the exam, greater consideration should be given to the impact that shorter response time lengths may have.

The inclusion of a speaking section in the latest version of the TOEFL test is a great step forward. Now that the test is in use, more research is necessary to ensure that the TOEFL iBT test provides a positive experience for test takers, both on and in the lead up to the test day.

Bio Data

Kristen Sullivan is a lecturer at Shimonoseki City University and cowriter of *Impact Conversation 1 & 2*. She is interested in the teaching, learning, and assessment of speaking, as well as interactions between language learner identity and language use. <kris@shimonoseki-cu.ac.jp>

References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). *TOEFL 2000 speaking framework: A working paper*. TOEFL Monograph Series, Report No. 20. Princeton, NJ: Educational Testing Service.
- Chapelle, 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. (2007). *The official guide to the new TOEFL iBT*. New York: McGraw-Hill.
- Educational Testing Service. (2009). *The official guide to the TOEFL test* (3rd ed.). New York: McGraw-Hill.
- Enright, M. K., Bridgeman, B., Eignor, D., Lee, Y., & Powers, D. E. (2008). Prototyping new assessment tasks. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 97-144). New York: Routledge.
- Hughes, R. (2002). *Teaching and researching speaking*. London: Pearson Education.
- Swain, M. (1984). Large-scale communicative language testing: A case study. In S. J. Savignon (Ed.), *Initiatives in communicative language teaching: A book of readings* (pp. 185-201). Reading, MA: Addison-Wesley.
- Wigglesworth, G., & Elder, C. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly*, 7, 1-24.

Yoshida, R. (2013, April 8). LDP panel binds TOEFL to degrees. *The Japan Times*. Retrieved from <http://www.japantimes.co.jp>