



APPENDIX B

Using Timing Data to Investigate Non-Response in the TIMSS 2019 Problem Solving and Inquiry (PSI) Tasks

Overview

Timing data collected as part of eTIMSS 2019 can be used to learn more about the patterns of non-response and block position effects described in Appendix A. Considering the relatively large percentages of students not reaching all of the items in the assessment eBooklets containing Problem Solving and Inquiry (PSI) tasks, it was clear that not all students had an opportunity to answer all of the items. However, reviewing the response data for the PSI tasks together with timing data indicated that instead of running out of time, some students stopped responding to items with plenty of time remaining, perhaps through fatigue or lack of motivation.

To explore this possibility, an investigation was conducted using measures derived from event log data collected as part of eTIMSS 2019. Event log data provide a comprehensive sequence of students' interactions with the computer-based assessment. A timestamp (in milliseconds) was saved for each student-object interaction, providing a full history of what the student clicked, entered, and selected. From the log data, information could be derived about students' test-taking behaviors, including response time, item visit and revisit behavior, and response revisions. Analyzing this information alongside the response data provided additional insights into student patterns of non-response.

Two phases of analysis were conducted: an item-level analysis and a student-level analysis. Analyses were conducted separately for each of fourth and eighth grades and for mathematics and science at the booklet level, focusing on the last item of each subject part (session). First, examining average timing measures for each PSI item showed that some students who did not reach all items actually had time remaining after giving their last response, but stopped responding to subsequent items. Then, to determine how many students stopped responding before time was up versus ran out of time, students who did not reach all items were classified into groups based on the time they gave their last response as a proxy for their last meaningful interaction with the items. The results for PSI students were compared to the results for students who took "regular" (non-PSI) eTIMSS items and indicated that the stopping behavior was much more common in the PSI tasks, particularly in mathematics, and was associated with lower

performance on the PSIs. In mathematics, relatively more students stopped responding than ran out of time. In science, smaller percentages of students in each of the two groups were about equal.

Definition of Not Reached Items

The TIMSS definition of “Not Reached” assumes that students progress through assessment booklets, or eBooklets, in sequential order. Following TIMSS’ standard data cleaning procedures developed for paper-based administration, if a student omitted two items in a row and all subsequent items in the booklet half were also blank, the second omitted item and all subsequent were coded in the data as “Not Reached.” In the paper-based environment, there is no way to know if the student was working on the first omitted item, and so it was considered to be omitted rather than not reached. However, with the new information available through event log data, it was possible to determine whether a student actually visited an item screen, the time they arrived, and how long they spent. The analyses conducted for this Appendix take advantage of this information to make informed inferences about whether students did not reach all items because they ran out of time or because they stopped responding, perhaps because of fatigue or lack of motivation.

Item-Level Analysis

The item-level timing analysis revealed evidence that there were at least some students who stopped responding to PSI items sometime before the end of each session or subject part. This was true at both fourth and eighth grades and particularly in mathematics. Students were given the same amount of time to complete each of two sessions—36 minutes at the fourth grade and 45 minutes at the eighth grade. Students were given two blocks of mathematics PSI items in one session and two blocks of science PSI items in the other session, with a 15-minute break in between.

The item-level analysis indicated for each PSI item the percentage of students who were coded as not reaching the item, but who had a record of arriving on the screen containing the item. Timing averages for these students were compared to averages for students who did reach the item. This included 1) the average time that students arrived on the item screen (in minutes from 0—the start of the session), and 2) the average time that students spent on the screen (in minutes).

Exhibit B.1 presents the results for the last mathematics item in each PSI booklet at the fourth grade. As described in Appendix A, eBooklet 15 included two blocks of mathematics PSIs in the first half of the assessment—*Penguins + Robots-4*, followed by *School Party*. eBooklet 16 had *School Party* then *Penguins + Robots-4* in the second half. Fourth grade students had 36 minutes to complete the two PSI blocks.

On average across countries, students who reached the last *School Party* item in eBooklet 15 arrived at 27.3 minutes, leaving 8.7 minutes remaining until the end of the session (at 36 minutes).

Students who reached this item spent 1.84 minutes responding to the item, on average. Although 39 percent of students had the last item coded as “not reached,” almost half of these students (16%) visited the screen at least once. The 16 percent of students with the item not reached arrived at 30.5 minutes with 5.5 minutes remaining, suggesting that these students arrived with plenty of time left, but did not respond. The small amount of time spent by the not-reached students (0.75 minutes) could be due to logging out of the test early or going back through previous screens.

Exhibit B.1: Timing Averages for the Last Mathematics PSI Item by eBooklet—Grade 4

eBooklet	Last Item Reached		Last Item Not Reached			
	Arrival Time (minutes)	Time Spent (minutes)	Total Percent	Percent Visited Screen	Arrival Time (minutes)	Time Spent (minutes)
eBooklet 15 (Positions 1 & 2)	27.3	1.84	39%	16%	30.5	0.75
eBooklet 16 (Positions 3 & 4)	25.1	2.48	30%	17%	25.2	1.50

Fourth grade students had 36 minutes to complete two blocks of mathematics PSIs. Timing information is limited to the screen level. Percentages are at the item level.

The results for eBooklet 16 were similar, but with relatively fewer students failing to reach the last item (30% compared to 39% in eBooklet 15). On average, students who reached the last *Robots-4* item arrived at 25.1 minutes with 10.9 minutes remaining, and spent 2.48 minutes on the screen. More than half of the students who did not reach the item (17%) actually had been recorded as arriving on the screen at 25.2 minutes with 10.8 minutes remaining, on average. These students had much more time remaining than the average time spent by students who responded to the item (2.48 minutes).

Exhibit B.2 presents the results for fourth grade science. Fewer fourth grade students did not reach all the science items than did not reach all the mathematics items—19 percent in eBooklet 15 and 28 percent in eBooklet 16 (compared to 39% and 30% for mathematics, respectively). More students failed to reach all items (or did not respond to all items) when *Sugar Experiment* followed by *Farm Investigation* were in the first session than when *Farm Investigation* followed by *Sugar Experiment* were in the second session (28% in eBooklet 16 vs. 19% in eBooklet 15).

Exhibit B.2: Timing Averages for the Last Science PSI Item by eBooklet—Grade 4

eBooklet	Last Item Reached		Last Item Not Reached			
	Arrival Time (minutes)	Time Spent (minutes)	Total Percent	Percent Visited Screen	Arrival Time (minutes)	Time Spent (minutes)
eBooklet 15 (Positions 3 & 4)	25.6	0.59	19%	7%	24.8	0.13
eBooklet 16 (Positions 1 & 2)	27.3	1.19	28%	3%	30.8	0.13

Fourth grade students had 36 minutes to complete two blocks of science PSIs. Timing information is limited to the screen level. Percentages are at the item level.

As reported in Exhibit B.2, students who reached the last science item in eBooklet 15 arrived at 25.6 minutes and in eBooklet 16 arrived at 27.3 minutes, on average. There were small percentages of students with the last item coded as “not reached” that actually visited the screen—7 percent in eBooklet 15 and 3 percent in eBooklet 16. However, given that students who reached the last item spent only 0.59–1.19 minutes on average, the students in question seem to have had ample time remaining to respond, arriving with 11.2 minutes remaining (eBooklet 15) and 5.2 minutes remaining (eBooklet 16), on average.

At the eighth grade, students had 45 minutes to complete each half of their booklets, but had more items to answer compared to fourth grade. Exhibit B.3 presents the results for the last eighth grade mathematics PSI items in eBooklets 15 and 16, respectively. In both booklets, the majority of students with not-reached codes for the last item had record of visiting the item screen—9 percent in eBooklet 15 (compared to 14% total not reached) and 18 percent in eBooklet 16 (compared to 21% total not reached), on average. These students arrived with 14.0–16.3 minutes remaining on the clock, on average. Students who reached the item spent 1.99–2.70 minutes responding on average, suggesting there was enough time remaining for at least some of the not-reached group to respond.

Exhibit B.3: Timing Averages for the Last Mathematics PSI Item by eBooklet—Grade 8

eBooklet	Last Item Reached		Last Item Not Reached			
	Arrival Time (minutes)	Time Spent (minutes)	Total Percent	Percent Visited Screen	Arrival Time (minutes)	Time Spent (minutes)
eBooklet 15 (Positions 1 & 2)	30.2	1.99	14%	9%	31.0	0.48
eBooklet 16 (Positions 3 & 4)	27.6	2.70	21%	18%	28.7	1.39

Eighth grade students had 45 minutes to complete two blocks of mathematics PSIs. Timing information is limited to the screen level. Percentages are at the item level.

Similar to the results at the fourth grade, the eighth grade mathematics results showed evidence that the ordering of the PSIs within the booklet had an effect on average completion rates. In eBooklet 16 with *Dinosaur Speed* followed by *Building + Robots-8* in the second half, more students failed to reach all items than in eBooklet 15 when *Building + Robots-8* followed by *Dinosaur Speed* were in the first half (21% vs. 14%).

Based on the results for eighth grade science (Exhibit B.4), it could be said that the eighth grade science PSIs were the most successful in keeping students engaged. In eBooklet 15, only 5 percent of students did not reach the last item, and in eBooklet 16 only 8 percent did not reach the last item. Only 3 percent of not-reached students in each booklet had record of visiting the last item screen.

Exhibit B.4: Timing Averages for the Last Science PSI Item by eBooklet—Grade 8

eBooklet	Last Item Reached		Last Item Not Reached			
	Arrival Time (minutes)	Time Spent (minutes)	Total Percent	Percent Visited Screen	Arrival Time (minutes)	Time Spent (minutes)
eBooklet 15 (Positions 3 & 4)	27.9	0.81	5%	3%	23.4	0.10
eBooklet 16 (Positions 1 & 2)	30.4	1.69	8%	3%	33.3	0.28

Eighth grade students had 45 minutes to complete two blocks of science PSIs. Timing information is limited to the screen level. Percentages are at the item level.

Student-Level Analysis

The item-level analysis made evident that at least some of the students who did not reach all items had enough time remaining, but stopped responding before the time expired. The next step involved determining the relative proportions of students who exhibited this stopping behavior versus running out of time. Toward this end, a time was derived for each student to indicate when their last response was given (or revised) in each booklet half (minutes from 0—the start of the session). This measure for “time of last response” served as an approximation of the time that students last meaningfully interacted with an item during the subject session.

Unfortunately, it was not possible to determine the precise time that students logged out of the test session. Therefore, it was necessary to implement a decision rule for when students had finished work on the assessment. This analysis assumes that among students who did not reach all items, those who remained active and gave a response within 30 seconds of the maximum allotted time (36 minutes at the fourth grade; 45 minutes at the eighth grade) ran out of time. On the other hand, those who did not interact with any item within 30 seconds of the time limit were

assumed to have stopped responding. This 30 second cutoff was chosen based on an analysis of the distribution of time of last response across countries.^{1,2}

At the fourth and eighth grades and for each subject, all students were classified according to the procedure described above, including all PSI students as well as students who took regular eTIMSS eBooks 1–14. First, students who reached all items were classified as “Reached All Items.” Among the students remaining, those who gave their last response within 30 seconds of the time limit (more than 35.5 minutes at the fourth grade; more than 44.5 minutes at the eighth grade) were classified as “Ran Out of Time.” The students who did not interact with any item within 30 seconds of the time limit were classified as “Stopped Responding.”

Exhibit B.5 presents the results for fourth grade mathematics. On average across countries, only 3 percent of students who took regular eTIMSS “Ran Out of Time” and 5 percent “Stopped Responding,” with the majority (92%) reaching all items. The 5 percent who stopped responding gave their last response at 29.6 minutes, with 6.4 minutes remaining, on average.

Exhibit B.5: Student Response Type Classifications for Mathematics—Grade 4

Student Response Group	Percent of Students	Time of Last Response (minutes)	Percent of Items Correct	Percent of Items Not Reached
Regular eTIMSS Mathematics				
1. Reached All Items	92%	25.8	49%	0%
2. Ran Out of Time	3%	35.9	40%	16%
3. Stopped Responding	5%	29.6	36%	15%
Total		26.2	49%	1%
PSI Mathematics				
1. Reached All Items	66%	29.0	40%	0%
2. Ran Out of Time	14%	35.9	43%	15%
3. Stopped Responding	21%	29.7	36%	13%
Total		30.1	40%	5%

All statistics were computed at the student level by country, then averaged across countries. Because of rounding some results may appear inconsistent.

The results for students who took PSI booklets tell a different story, with fewer fourth grade students having “Reached All Items” (66% of PSI students compared to 92% of regular students). Among PSI students, a higher percentage of students were classified as “Stopped Responding” compared to “Ran Out of Time”—21 percent versus 14 percent. On average across countries, PSI students who reached all mathematics items gave their last response at 29.0 minutes, with 7.0 minutes remaining in the session. Similar to the students who took the regular eTIMSS items, PSI students who stopped responding gave their last response at 29.7 minutes, not interacting with any item for more than 6 minutes, on average. As could be expected, these students had the

lowest performance (36% of items correct) compared to students who reached all items and students who ran out of time (40–43% of items correct).

Results for fourth grade science are shown in Exhibit B.6. Similar to mathematics, the majority of students who took regular eTIMSS (93%) reached all items with an average time of 25.5 minutes. On the other hand, just 76 percent of PSI students reached all science items, and with an average time of 28.1 minutes. In contrast with the mathematics results which showed that relatively more students stopped responding than ran out of time (21% vs. 14%), science had about equal amounts of students in the two categories—13 percent and 11 percent, respectively. On average, students who stopped responding gave their last response at 30.6 minutes, with more than 5 minutes remaining. Students who stopped responding had the lowest average performance compared to the other two groups (35% of items correct compared to 44–45% of items correct).

Exhibit B.6: Student Response Type Classifications for Science—Grade 4

Student Response Group	Percent of Students	Time of Last Response (minutes)	Percent of Items Correct	Percent of Items Not Reached
Regular eTIMSS Science				
1. Reached All Items	93%	25.5	53%	0%
2. Ran Out of Time	3%	35.8	43%	18%
3. Stopped Responding	4%	31.8	40%	20%
Total		25.9	52%	1%
PSI Science				
1. Reached All Items	76%	28.1	45%	0%
2. Ran Out of Time	11%	35.9	44%	19%
3. Stopped Responding	13%	30.6	35%	22%
Total		29.2	44%	5%

All statistics were computed at the student level by country, then averaged across countries. Because of rounding some results may appear inconsistent.

The results for eighth grade mathematics in Exhibit B.7 show the majority of students who took regular eTIMSS reaching all items (94%) and relatively fewer PSI students (83%) doing so. Similar to the mathematics results at the fourth grade, more eighth grade PSI students stopped responding than ran out of time (14% vs. 3%). PSI students who reached all items gave their last response at 34 minutes with 11 minutes remaining, which is similar to the regular eTIMSS students who finished at 33.6 minutes, on average. In comparison, PSI students who stopped responding last interacted with an item two minutes earlier, at 32.1 minutes with 13.9 minutes remaining, on average.

Exhibit B.7: Student Response Type Classifications for Mathematics—Grade 8

Student Response Group	Percent of Students	Time of Last Response (minutes)	Percent of Items Correct	Percent of Items Not Reached
Regular eTIMSS Mathematics				
1. Reached All Items	94%	33.6	43%	0%
2. Ran Out of Time	2%	44.9	35%	14%
3. Stopped Responding	4%	35.7	32%	16%
Total		33.9	42%	1%
PSI Mathematics				
1. Reached All Items	83%	34.0	29%	0%
2. Ran Out of Time	3%	44.9	32%	15%
3. Stopped Responding	14%	32.1	25%	14%
Total		34.1	29%	3%

All statistics were computed at the student level by country, then averaged across countries. Because of rounding some results may appear inconsistent.

The results for eighth grade science (Exhibit B.8) were similar between regular eTIMSS students and PSI students, with the majority of students reaching all items in both groups (97% and 94%, respectively). Among PSI students, on average, the 4 percent who stopped responding had the lowest performance on the PSI tasks, with students who reached all items and ran out of time answering at least 10 percent more items correct than students who stopped responding (41–44% of items correct vs. 30% of items correct).

Exhibit B.8: Student Response Type Classifications for Science—Grade 8

Student Response Group	Percent of Students	Time of Last Response (minutes)	Percent of Items Correct	Percent of Items Not Reached
Regular eTIMSS Science				
1. Reached All Items	97%	31.1	47%	0%
2. Ran Out of Time	1%	–	–	–
3. Stopped Responding	2%	30.4	31%	17%
Total		31.2	47%	0%
PSI Science				
1. Reached All Items	94%	32.5	44%	0%
2. Ran Out of Time	3%	44.9	41%	15%
3. Stopped Responding	4%	31.8	30%	18%
Total		32.7	44%	1%

All statistics were computed at the student level by country, then averaged across countries.

A dash (–) indicates comparable data not available. Because of rounding some results may appear inconsistent.

Notes

- ¹ Soland, J., Kuhfield, M., & Rios, J. (2021). Comparing different response time threshold setting methods to detect low effort on a large-scale assessment. *Large-scale Assessments in Education*, 9(8). <https://doi.org/10.1186/s40536-021-00100-w>
- ² Ulitzsch, E., von Davier, M., & Pohl, S. (2019). A multiprocess item response model for not-reached items due to time limits and quitting. *Educational and Psychological Measurement*, 80(3), 522–547. <https://journals.sagepub.com/doi/full/10.1177/0013164419878241>