IES Institute of Education Sciences

# Use NAEP Process and Performance Data to Examine the Impact of Extended Time Accommodation

**Xin Wei, SRI International,**
**Susu Zhang, University of University of Illinois, Urbana-Champaign**
**Jihong Zhang, University of Iowa**

**Annual IES Principal Investigators Meeting**
**January 25-27, 2022**
***Advancing Equity and Inclusion in the Education Sciences***

# Introduction

- Extended time accommodation is the most common testing accommodation, yet the effects of extended time accommodation are not well-understood
- Our study uses process data and performance data for the digital NAEP assessment to examine the performance and behavior differences between students with learning disability (LD) who received extended time accommodation and their peers with LD who did not receive extended time accommodation.

# Extended Time Accommodation

- The objective of educational assessments should be to accurately reflect learners' knowledge and capabilities in a content area rather than their disabilities.
- Federal legislation mandates the inclusion of students with disabilities in the administration of large-scale academic assessments and permits their use of accommodations
- Test accommodations allow SWDs to demonstrate their ability and knowledge in an area without the interference of their disabilities and to create equal opportunities for SWDs

# Effectiveness of Extended Time Accommodation

- A meta-analysis found that extended time accommodation is associated with improved test scores for both students with LD and their typically developing (TD) peers (Gregg & Nelson, 2012).
- However, students with LD with an extended time accommodation still underperformed their TD peers who did not receive extended time (Gregg & Nelson, 2012).
- Test accommodations has social emotional benefits (Wadley & Liljequist, 2013).

# Limitation of Previous Studies

- Small, nonrepresentative, convenience samples
- Very few within students with LD group comparison of the effect of extended time accommodation
- No studies have examined the impact of extended time accommodation on test-taking behavior (time spent, number of visits, and number of actions on each item)

# Research Questions

- Do students with LD in the extended time condition spend more time, exhibit higher number of revisits, perform more actions, and score higher than their peers with LD who did not receive the extended time accommodation?

# Sample

- 600 students with LD who received extended-time accommodation (90 minutes) vs. 930 students with LD who did not receive extended time accommodation (30 minutes)
- Both completed one NAEP math assessment block of 15 items

# Measures

- NAEP is the nationally administered low-stakes test that represent student achievement in the U.S.
- 15 math items (multiple choice, constructed response, or matching items)
- Skipping and revisiting are allowed

# Variables

- School staff filled out a questionnaire for each student with disabilities who were sampled to participate in the NAEP study (NAEP, 2020). This disability questionnaire collected information about a student's disability category, whether this student participated in the state assessment with accommodations (and what type of accommodations should be provided), without accommodations, or met the participation criteria for the state alternate assessment.
- NAEP assessment followed test accommodation decision based on the information collected in the questionnaire.

# Variables

- Item-level performance
- Item response time and total response time
- Item-level total number of visits
- Item-level number of actions
- Demographic characteristics

# Statistics

- Standardized mean difference (SMD) test to perform significant testing for differential item functioning (DIF), differential item response time (DRT), differential action sequence length (DSL), differential number of visits (DNV)
- Matching variable was initial performance, i.e., total score on the first five questions.

# Sample Characteristics

| Demographics and math performance | 30 minutes (n=600) | | 90 minutes (n= 930) | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Male | 64.27% | 0.48 | 60.37% | 0.49 |
| Age | 14.60 | 0.60 | 14.63 | 0.596 |
| White | 52.42% | 0.50 | 47.26% | 0.50 |
| African American | 13.52% | 0.34 | 20.41% | 0.40 |
| Hispanic | 21.37% | 0.41 | 26.32% | 0.44 |
| Other | 12.69% | 0.33 | 6.02% | 0.24 |
| Math total points on all 15 items | 5.28 | 2.85 | 5.03 | 2.87 |
| Math total points on first 5 items | 3.65 | 3.16 | 3.41 | 3.08 |
| Math total points on first 5 items separated into 5 groups | | | | |
| 0-1 | 16.47% | .37 | 20.67% | .41 |
| 2-3 | 41.51% | .49 | 41.12% | .49 |
| 4-5 | 31.76% | .47 | 30.03% | .46 |
| 6-7 | 10.25% | .30 | 8.18% | .27 |
| Total response time on 15 items in seconds | 1376.60 | 388.34 | 1550.92 | 734.08 |

# Differential Response Times

| Item | Mean (SD) of Log (Resp. Time + 1) | | DRT | |
|---|---|---|---|---|
| | 30 minutes | 90 minutes | SMD | p |
| 1 | 4.96 (4.84) | 5.11 (5.09) | -0.1** | 0.006 |
| 2 | 5.41 (4.96) | 5.42 (4.95) | 0.02 | 0.74 |
| 3 | 5.61 (5.2) | 5.65 (5.26) | -0.006 | 0.419 |
| 4 | 5.46 (5.06) | 5.57 (5.35) | -0.03 | 0.175 |
| 5 | 5.62 (5.26) | 5.68 (5.43) | -0.03 | 0.202 |
| 6 | 5.71 (5.15) | 5.74 (5.31) | -0.003 | 0.471 |
| 7 | 5.64 (5.53) | 5.65 (5.54) | -0.02 | 0.338 |
| 8 | 5.12 (4.83) | 5.24 (5.24) | -0.04 | 0.189 |
| 9 | 5.33 (5.06) | 5.39 (5.16) | -0.02 | 0.317 |
| 10 | 5.36 (4.93) | 5.5 (5.43) | -0.06 | 0.096 |
| 11 | 6.05 (5.54) | 6.16 (5.91) | -0.068 | 0.107 |
| 12 | 5.14 (4.68) | 5.27 (5.54) | -0.11** | 0.007 |
| 13 | 6.01 (5.47) | 6.22 (5.9) | -0.23*** | 0.001 |
| 14 | 5.3 (5.03) | 5.59 (5.48) | -0.42*** | 0.001 |

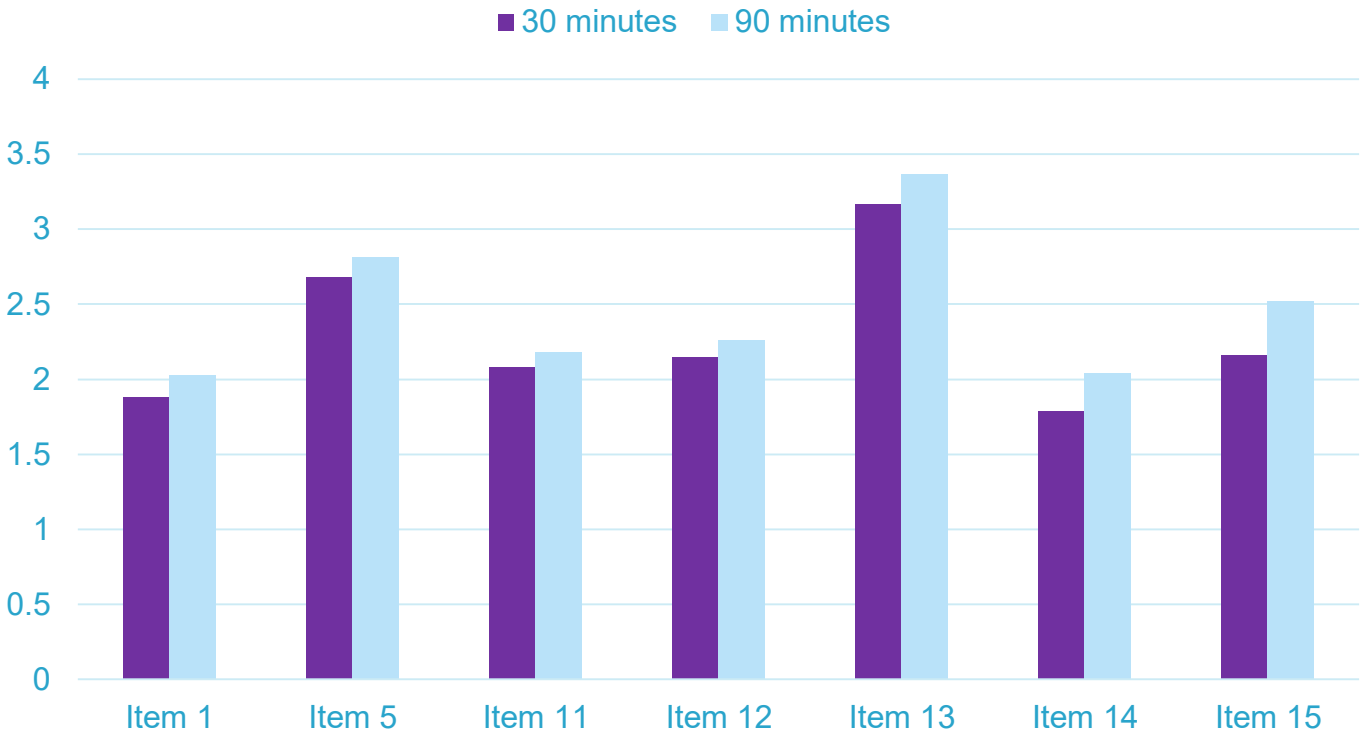
# Log of Response Time

# Differential Sequence Length

| Item | Mean (SD) of Log (Ave. Seq. Len. + 1) | | DSL | |
|---|---|---|---|---|
| | 30 minutes | 90 minutes | SMD | p |
| 1 | 1.88 (2.53) | 2.03 (2.69) | -0.1* | 0.033 |
| 2 | 2.34 (2.14) | 2.4 (2.41) | -0.02 | 0.288 |
| 3 | 2.79 (3.22) | 2.85 (3.34) | -0.04 | 0.279 |
| 4 | 2.02 (2.52) | 2.02 (2.53) | -0.02 | 0.32 |
| 5 | 2.68 (2.82) | 2.81 (3.08) | -0.09* | 0.012 |
| 6 | 2.46 (2.26) | 2.5 (2.53) | -0.01 | 0.273 |
| 7 | 2.69 (3.43) | 2.55 (3.13) | -0.03 | 0.333 |
| 8 | 1.95 (2.35) | 1.92 (2.44) | 0.04 | 0.834 |
| 9 | 2.32 (2.77) | 2.34 (2.54) | -0.04 | 0.119 |
| 10 | 2.42 (2.51) | 2.45 (2.65) | -0.02 | 0.343 |
| 11 | 2.08 (2.29) | 2.18 (2.52) | -0.07* | 0.031 |
| 12 | 2.15 (2.25) | 2.26 (2.53) | -0.08* | 0.051 |
| 13 | 3.17 (2.86) | 3.37 (3.07) | -0.28*** | 0.001 |
| 14 | 1.79 (2.2) | 2.04 (2.96) | -0.24*** | 0.001 |

Institute of Education Sciences

# Log of Sequence Length

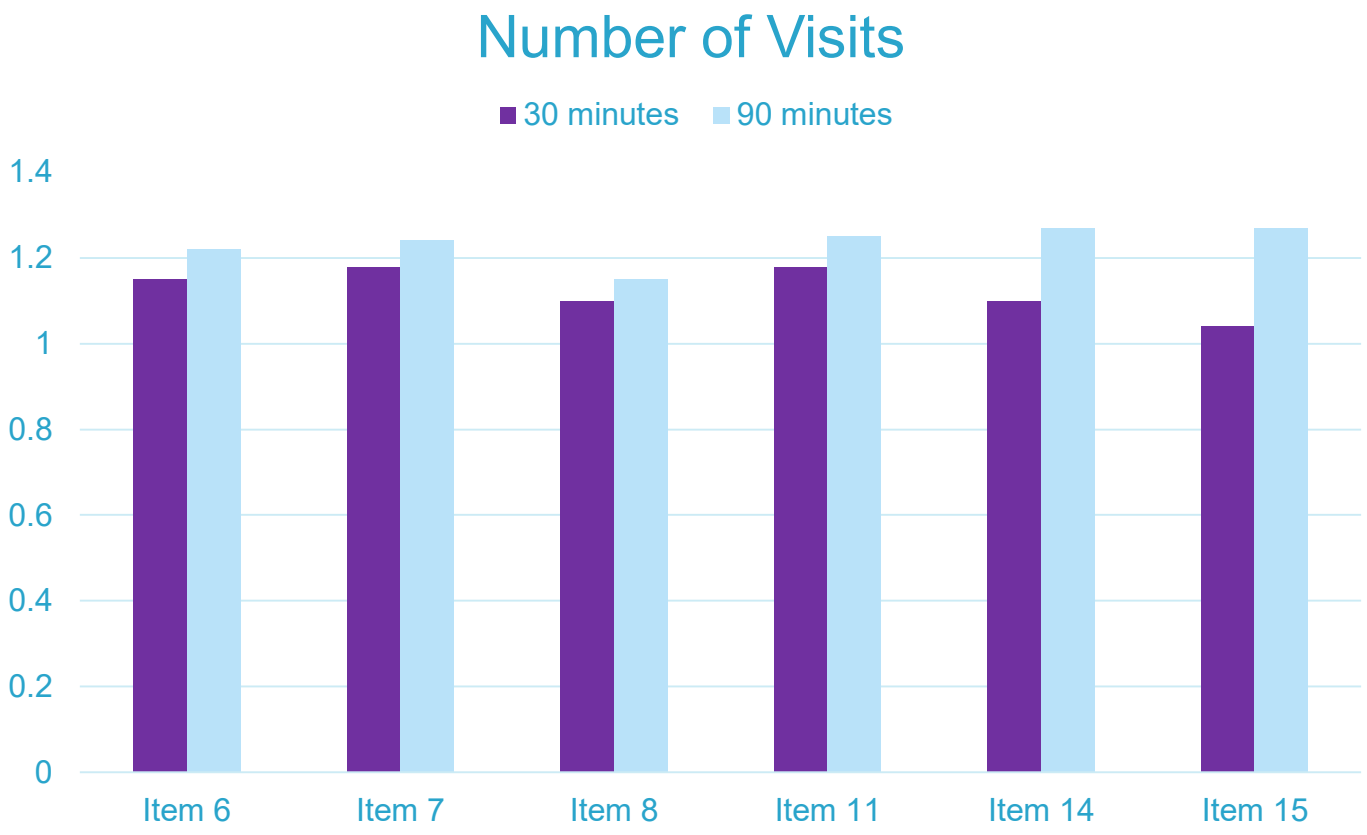


Institute of Education Sciences

# Differential Number of Visits

| Item | Mean (SD) of number of visits | | DNV | |
|---|---|---|---|---|
| | 30 minutes | 90 minutes | SMD | p |
| 1 | 1.14 (0.48) | 1.17 (0.51) | -0.03 | 0.134 |
| 2 | 1.12 (0.46) | 1.14 (0.41) | -0.01 | 0.259 |
| 3 | 1.13 (0.43) | 1.14 (0.51) | -0.01 | 0.322 |
| 4 | 1.14 (0.42) | 1.12 (0.44) | 0.02 | 0.85 |
| 5 | 1.52 (1.19) | 1.63 (1.52) | -0.09 | 0.104 |
| 6 | 1.15 (0.45) | 1.22 (0.62) | -0.06* | 0.017 |
| 7 | 1.18 (0.54) | 1.24 (0.62) | -0.06* | 0.028 |
| 8 | 1.1 (0.41) | 1.15 (0.48) | -0.04* | 0.049 |
| 9 | 1.08 (0.3) | 1.11 (0.38) | -0.03 | 0.068 |
| 10 | 1.16 (0.81) | 1.16 (0.47) | -0.002 | 0.483 |
| 11 | 1.18 (0.51) | 1.25 (0.61) | -0.06* | 0.02 |
| 12 | 1.13 (0.48) | 1.15 (0.48) | -0.02 | 0.206 |
| 13 | 1.6 (1.09) | 1.66 (1.03) | -0.05 | 0.182 |

Institute of Education Sciences



# Differential Item Functioning

| Item | Mean (SD) of item score | | DIF | |
|---|---|---|---|---|
| | 30 minutes | 90 minutes | SMD | p |
| 1 | 0.31 (0.46) | 0.29 (0.46) | -0.01 | 0.253 |
| 2 | 0.89 (0.32) | 0.87 (0.33) | -0.001 | 0.457 |
| 3 | 0.22 (0.41) | 0.17 (0.38) | 0.03 | 0.946 |
| 4 | 0.79 (0.79) | 0.71 (0.80) | 0.02 | 0.711 |
| 5 | 1.05 (0.82) | 0.98 (0.83) | -0.001 | 0.468 |
| 6 | 0.16 (0.37) | 0.14 (0.35) | 0.003 | 0.55 |
| 7 | 0.14 (0.35) | 0.12 (0.33) | 0.02 | 0.862 |
| 8 | 0.24 (0.43) | 0.23 (0.42) | 0.004 | 0.579 |
| 9 | 0.43 (0.50) | 0.41 (0.49) | 0.004 | 0.586 |
| 10 | 0.21 (0.47) | 0.21 (0.46) | -0.01 | 0.357 |
| 11 | 0.04 (0.28) | 0.05 (0.30) | -0.01 | 0.278 |
| 12 | 0.07 (0.31) | 0.05 (0.27) | 0.01 | 0.734 |
| 13 | 0.40 (0.65) | 0.45 (0.67) | -0.07* | 0.03 |
| 14 | 0.30 (0.62) | 0.29 (0.61) | 0.006 | 0.585 |

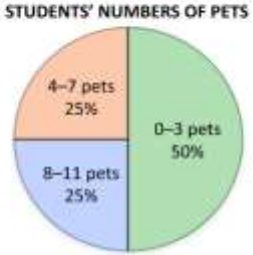
# Item 13

VH302334

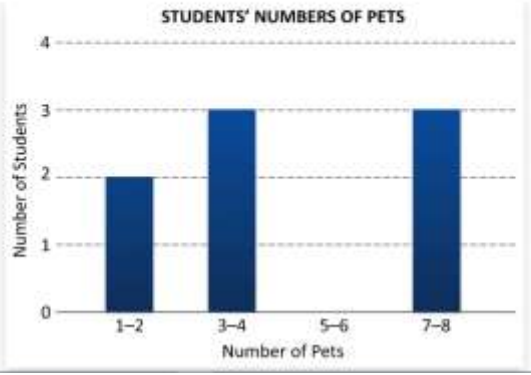
A teacher surveyed 8 students in a class to find out how many pets they each have.

Boyd and Jenny represented the results of the survey in two different ways.

Boyd made a circle graph (pie chart) to show the results.




Jenny made a bar graph to show the results.




Based on the circle graph, which of the following sets of data could be the results of the survey?

Select all the correct answers.

- A ☐ 0, 1, 2, 2, 3, 3, 5, 10
- B ☐ 0, 1, 2, 3, 5, 6, 9, 11
- C ☐ 0, 2, 2, 3, 4, 8, 9, 10
- D ☐ 1, 1, 1, 1, 4, 4, 8, 8
- E ☐ 1, 2, 4, 5, 6, 7, 8, 10

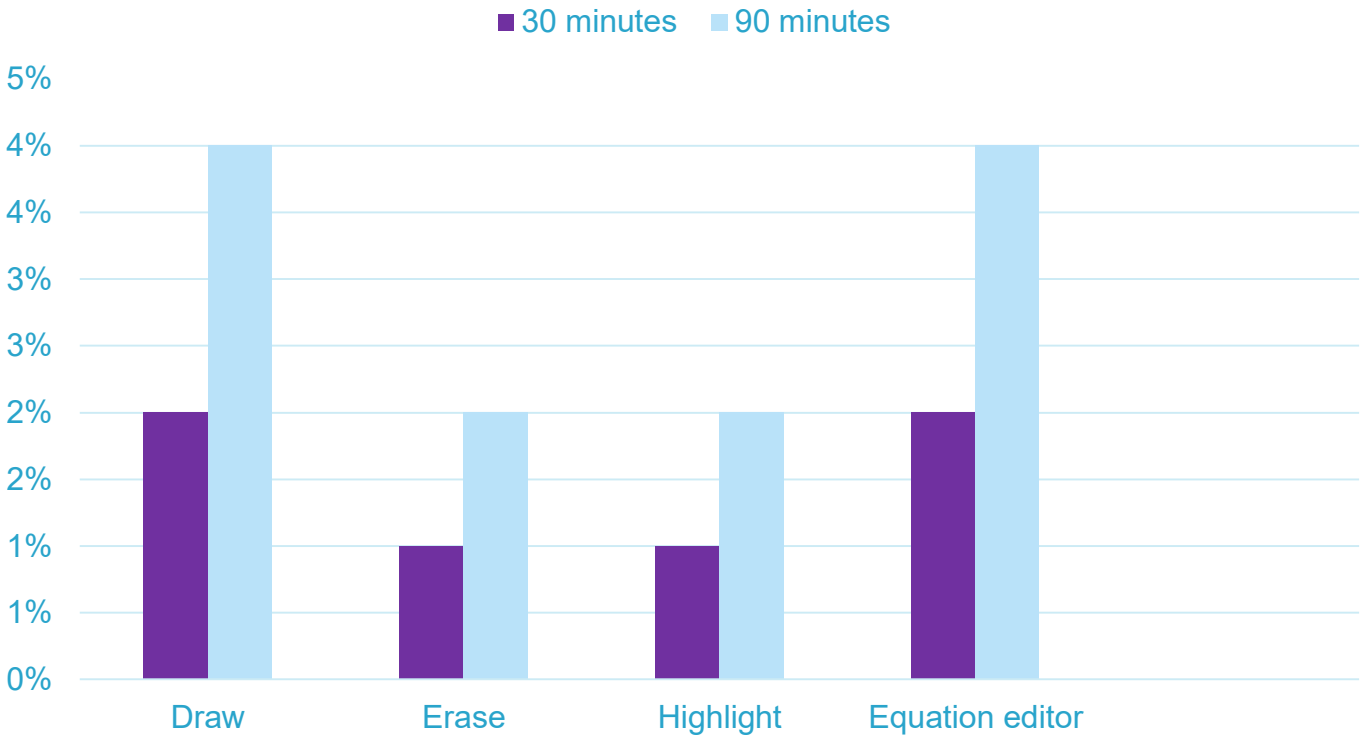
Clear Answer

Based on both Boyd's circle graph and Jenny's bar graph, what could be the results of the survey?

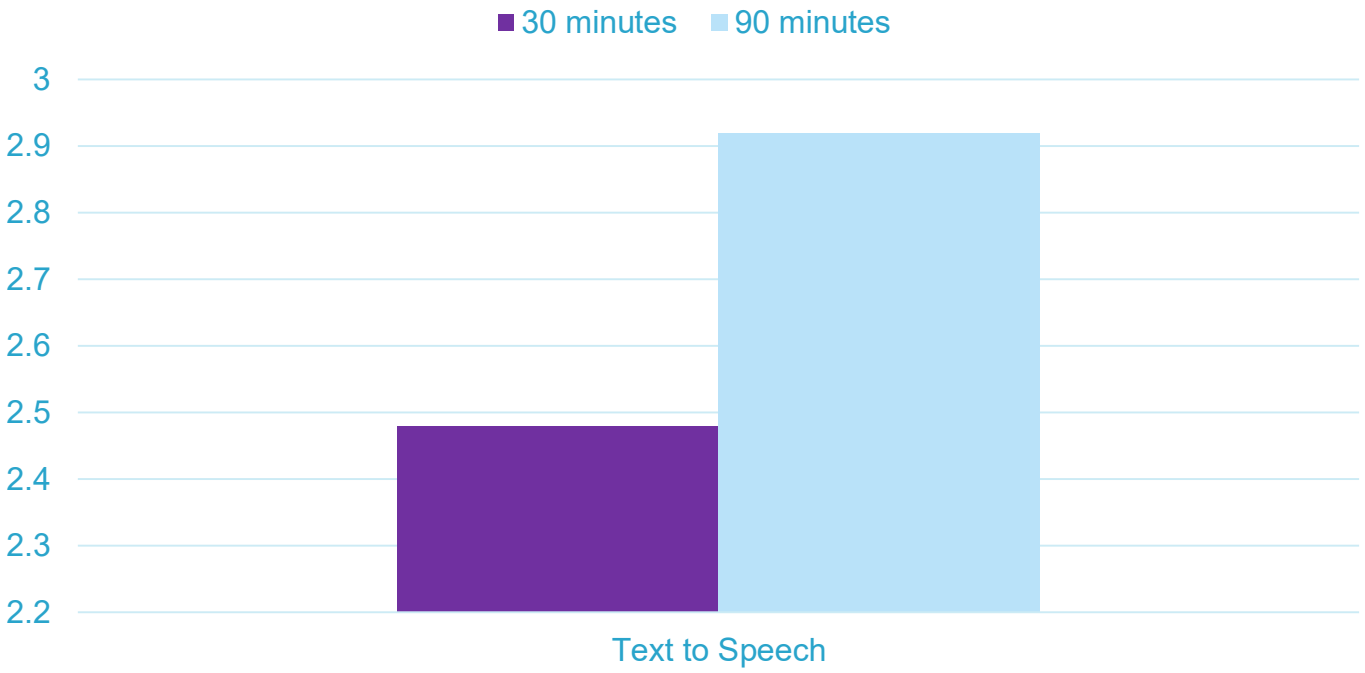
Enter a number in each box to show how many pets each student could have.

| | | | |
|---|---|---|---|
| | | | |
| | | | |

# UDL Tool Usage on Item 13



# Number of Times Used Text to Speech on Item 13



# Results

- Students with LDs in the extended time condition are more likely to be minority or female and scored lower on NAEP than students with LDs in the standard testing time condition
- Comparable students with LDs in the extended time accommodation group spent significantly more time on the Items 1, 12, 13, 14, and 15 than their peers with LDs in the standard test condition
- Comparable students with LDs in the extended time condition had a significantly higher number of actions than their peers tested under standard conditions on Items 1, 5, 11, 12, 13, 14, 15

# Results

- Comparable students with LDs in the extended time condition made more visits to Items 6, 7, 8, 11, 14, and 15 of the items than their peers tested under the standard condition
- One item, Item 13, was found to be biased in favor of students with LDs in the extended time accommodation condition; this indicates students with LDs in the extended time accommodation received a significantly higher average score on item 13 than their peers tested under the standard condition, given the same the performance on the first five items.

# Implications

- Extended time accommodation alleviates the speededness problem of timed assessment by providing extra time to students with LD to interact with the test items (particularly those at the end of test), come back to an item multiple times to work on that item, use universal design tools to facilitate their understanding, and try different strategies to solve the problems.
- The significant DIF on Item 13 is an indication that that item is not likely to measure intended abilities or constructs but instead secondary or irrelevant abilities (for example, speededness) which can jeopardize test validity and equity (Lu & Sireci, 2007).
- Receiving extra time might be particularly beneficial for students with LDs to work on medium level difficult items that are located toward the end of the test.

THANKS!

# Any questions?

You can find me at xin.wei@sri.com