IES Institute of Education Sciences

NAEP Mathematics Data for Students with Disabilities

Technical Working Group Summary | March 18, 2020

U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Special Education Research at IES

National Assessment of Educational Progress (NAEP) Mathematics Data for Students with Disabilities

Technical Working Group (TWG) Meeting

March 18, 2020

National Center for Special Education Research
Institute of Education Sciences
U.S. Department of Education
Washington, DC

Technical working group members reviewed and commented on the following meeting summary, and IES incorporated their corrections and edited the summary for clarity and consistency. The views expressed in this document reflect both individual and collective opinions of the meeting participants and are not necessarily those of the U.S. Department of Education.

Contents

Technical Working Group Members

<u>Invited Experts</u>

Jennifer Cain, NAEP State Coordinator
Minnesota Department of Education

Ruhan Circi, Director of Process Data Center
American Institutes for Research

Jodi Davenport, Deputy Director for the Science, Technology, Engineering, & Mathematics
WestEd

Lizanne DeStefano, Professor of Psychology and the Executive Director of the Center for Education Integrating Science, Mathematics, & Computing
Georgia Institute of Technology

Sidney D'Mello, Associate Professor
University of Colorado, Boulder

Anne Foegen, Professor and Director of Graduate Education in the School of Education
Iowa State University

Russell Gersten, Executive Director
Instructional Research Group

Leanne Ketterlin-Geller, Professor and Director of Research in Mathematics Education
Southern Methodist University

Andrew Krumm, Assistant Professor
University of Michigan Ann Arbor

Steven Ritter, Founder and Chief Scientist
Carnegie Learning

Rajiv Satsangi, Assistant Professor
George Mason University

Melina Uncapher, Assistant Professor and Director of Education at the UC San Francisco Neuroscape Center and Director of NewSchools' EF+Math Program
University of California San Francisco & EF+Math Program

Marcelo Worsley, Assistant Professor of Learning Sciences and Computer Science
Northwestern University

Institute of Education Sciences Participants

Sarah Brasiel, Education Research Analyst and Program Officer
National Center for Special Education Research

Peggy Carr, Associate Commissioner
National Center for Education Statistics

Daniel McGrath, Chief of the Assessment Division in the Reporting and Dissemination Branch
National Center for Education Statistics

Joan McLaughlin, Commissioner
National Center for Special Education Research

Mark Schneider, Director
Institute of Education Sciences

Emmanuel Sikali, Data Scientist
National Center for Education Statistics

Grady Wilburn, Research Scientist for the Assessment Division of NAEP
National Center for Education Statistics

Lynn Woodworth, Commissioner
National Center for Education Statistics

Background

The Institute of Education Sciences (IES) is working to release process data from the 2017 eighth grade mathematics NAEP. The process data include keystrokes and use of tools and accommodations by students taking the assessment. IES also plans to release NAEP survey data collected from learners, teachers, and schools. On March 18, 2020, the IES National Center for Special Education Research (NCSER) virtually convened a group of experts to discuss gaps in research on eighth grade mathematics for students with disabilities (SWD) that could be addressed with the NAEP process data and data science techniques for this research. Invited experts also provided recommendations for researchers interested in using the NAEP process data. The Appendix includes the agenda for the meeting and briefs written by the expert participants prior to the meeting. We recommend that the briefs be reviewed so that there is context for this summary of the meeting discussion. Some of the experts used PowerPoint slides during the meeting. These slides are available on the NCSER Technical Working Group Meeting Summaries website (https://ies.ed.gov/ncser/whatsnew/techworkinggroup/). This report summarizes the introductory remarks made by participants for each session as well as the discussions that followed. Recommendations to IES from each participant are also provided at the end of this summary.

Introduction

The meeting started with introductions from Mark Schneider, Director of IES, Joan McLaughlin, Commissioner of NCSER, Lynn Woodworth, Commissioner of the National Center for Education Sciences (NCES), and Peggy Carr, Associate Commissioner of NCES. These speakers emphasized the need to better understand how to improve instruction and assessment for SWD, especially given historically low performance on the NAEP math assessment. The NAEP process data provide a unique opportunity to look deeply at how SWD take tests, process information, and answer assessment questions.

Lynn Woodworth shared the process they have been going through at NCES prior to the release of these data. They held a panel of process data experts who provided recommendations for file structure and data layout, as well as what should be and should not be included for privacy reasons. NCES also convened a panel of education researchers who provided input on variables everyone will need to compute to be included in the data set. This provided a foundation for this first release of the 2017 eighth grade NAEP mathematics assessment process dataset.

Peggy Car provided an example of the process data from a prompt on the 2016 writing assessment (See Figure 1). On the y-axis is every key stroke a student provided in their response to the writing prompt. The x-axis is time; students have 30 minutes to provide a response to the prompt. There are data for 4 students in this image (red, green, yellow, and blue). The dips represent places where the student went back and made changes to their essay. The student represented by red wrote the most in the 30 minutes, making some changes at the beginning of the essay. The student represented by green wrote a lot and towards the end of the 30 minutes started to make changes to various parts of the essay. The student represented by blue did not write as much and made a couple changes along the way. The student represented by yellow student did not write anything for the first 13 minutes and then started the essay. Most would guess that the student represented by red scored the highest, but all the students received the same score for their writing response. This example provides a window into the richness of the data and how it can be used to explore what students know and how they achieve certain responses.

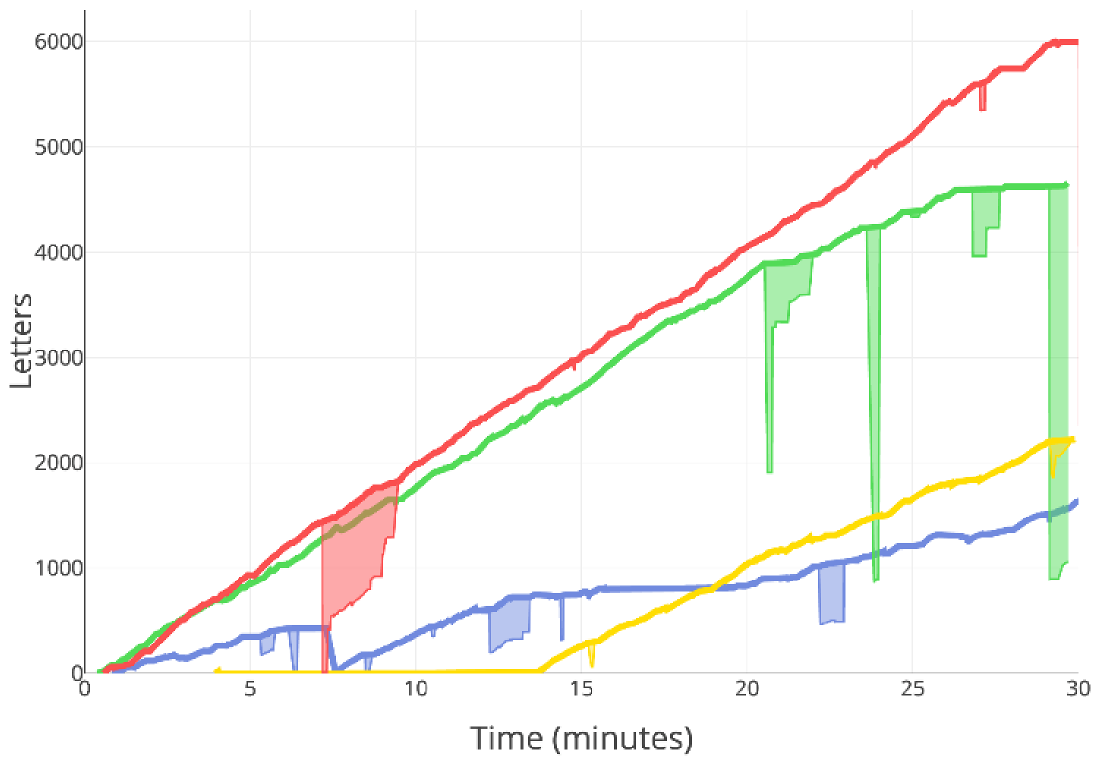


Figure 1. Example from 2016 NAEP Writing Assessment Process Data of Letters Typed Over Time

NAEP Assessment and Restricted-Use File(s)

Emmanuel Sikali, from NCES, provided a brief overview of the NAEP assessment design, the sample design, the data that will be made available to researchers, and inferences that can be made with this data. This information is not included in this summary because the full presentation will be provided as an on-demand webinar on the IES Research Funding Opportunities On-Demand Webinars website (https://ies.ed.gov/funding/webinars/index.asp).

NAEP Administration for Students with Disabilities

Grady Wilburn from NCES described the NAEP policies for SWD. An important NAEP policy encourages all jurisdictions to include at least 95% of all students and 85% of SWD and English language learners in the NAEP. Almost all jurisdictions are now meeting these targets. The two reasons for excluding students from the NAEP are (1) if students with severe cognitive disabilities take the alternative assessment with alternate achievement standards and the school or parent/guardian thinks they cannot meaningfully participate or (2) if students have an accommodation on their state test that the NAEP does not provide and the school or parent/guardian does not think they can meaningfully participate without that accommodation. Many SWD take the NAEP assessment, but the sample is not representative of all SWD because disability status is not a factor included in the sampling frame. For example, specialty schools serving only SWD are not in the sampling frame. Therefore, researchers must be careful generalizing to the whole population of SWD based on the NAEP data.

Jennifer Cain, NAEP State Coordinator for the Minnesota Department of Education, described the NAEP definition of SWD and the information that schools must provide about these students. NAEP policy defines SWD as those who have an individualized education program (IEP) for reasons other than being gifted or talented or who have a Section 504 plan that requires accommodations for the NAEP assessment. For each student with a disability, schools must record whether the student takes state or classroom assessments with accommodations. They must also provide information about accommodations each student needs for the NAEP assessment (choosing from a list of accommodations that are available on the NAEP and allowable by the state) or whether the student should not take the assessment. The goal is for students to participate on the NAEP in the way that they are most familiar with participating on other assessments.

Lizanne DeStefano of Georgia Institute of Technology discussed research questions that could be answered with the NAEP process and questionnaire data on SWD. NAEP data could be used to explore whether students use the accommodations that they were assigned, as a previous study by Dr. DeStefano indicated that some students do not use the accommodations that are available to them. In prior research she has also found that many SWD have not had experience with some of the accommodations in an instructional situation. For example, SWD could use a calculator on the state test, but many had not used a calculator during class instruction and did not use one during the test. Researchers could also look at the information provided in the NAEP student and teacher questionnaires to consider some of the differences in curriculum and instruction provided to SWD and students without disabilities. The NAEP process data could highlight the relationship between accommodations used, test-taking behavior, and item type to better understand how SWD approach the test, how that behavior is linked to their success or failure on an item, and whether there is value added by using particular accommodations. For example, keystroke and latency data and the amount of time on a question could help identify patterns of responding for SWD and could be linked to accommodations data.

Universal Design for Assessment (UDA) Features and Testing Accommodations

Marcelo Worsley of Northwestern University described the accommodations and UDA features (such as text to speech, zooming, and highlighting) available to SWD when they take the NAEP assessment as well as the limitations of those supports. First, the way questions are presented on the NAEP (for example, those that require students to interpret a graph) may present challenges for certain students with disabilities, such as those with visual impairment, that can't be overcome by accommodations or UDA features. Second, there may be some UDA features that are not provided but that students might need. For example, people who are sighted use a highlighter so that it is easier to go back to something or visually perceive the information, but there is not a parallel feature for students who do not have full vision. Researchers should consider the goal of the accommodation or UDA feature and the type of process data that could be used to validate whether it is achieving what was intended. Dr. Worsley also discussed the benefits of using the framework of ability-based design which focuses on designing systems to adapt to the abilities of the users and is consistent with the principles of UDA and several of the existing tools within the NAEP platform.

Leanne Ketterlin-Geller of Southern Methodist University noted that accessibility is the interaction point between the test taker (and their personal characteristics) and the test or the items themselves. UDA is meant to improve accessibility as it involves attending to a range of user characteristics when designing tests and items. Additional accommodations are applied to help support students who have personal characteristics that cannot be supported with UDA features. NAEP process data can give information on how helpful accommodations and UDA features are in decreasing the chance that a user characteristic may influence test scores beyond the student's knowledge of mathematics content. For example, these data could show how those using accommodations or UDA features perform in comparison with other students. The data could also show whether accommodations and UDA features have the desired effect of improving access to the NAEP test.

Accessibility features and accommodations are meant to even the playing field, but empirical studies are needed to determine how they are being used and whether they are doing what is intended. Researchers also need to keep in mind that several things may affect students' use of UDA features, including their awareness of these features, their ability to turn them on or off, their personal characteristics, and their disability/ability to use the features provided.

Discussion

If an accommodation is to have the hoped-for effect, the scores of SWD who use the accommodation(s) or UDA features should be similar to students without disabilities who are otherwise similar, across a number of problems. But both accommodations and item features are imperfect and this tension should be explored. Researchers could explore where the accommodations are not sufficient or students are not using the accommodations, as well as other instances where there are elements of the items themselves that seem to be a source of difficulty for SWD. Researchers could use the process data to understand the source of the disparities between students who have disabilities and students who do not.

Within the NAEP, the user has the option to select the UDA features they want to turn off or on for an item. It would be interesting to understand for what types of items students turn on those features and how it impacts their responding behavior. Researchers could also examine how these patterns differ for students with and without disabilities.

Student Cognition and Learning in Mathematics

Jodi Davenport of WestEd explained that taking a test involves several cognitive processes. These processes include attention, perception, memory, and reasoning. To show their ability, students must first understand the prompt. Students must engage their limited working memory to keep in mind the information they need to finish the task. Demand on working memory increases when students need to bring together pieces of information presented in different ways to answer a test question, such as a question that uses both text and a diagram. Once students understand the task and have an answer, they must provide the answer in the right format.

The format of certain questions may place more cognitive demands on students. For instance, if the problem requires students to respond to multiple prompts, students with attention difficulties might have trouble understanding how to parse out the different types of information in order to respond. The response type may also interact with a student's accommodations to make a task more complex. For example, items with a drag-and-drop feature may be more challenging for students using screen readers. It may be that certain item features make it easier for some students but more difficult for other students who need accommodations to respond.

Dr. Davenport emphasized that it is important for researchers to understand the context in which the data were collected and what the students were responding to when they generated the response. She also encourages research using the NAEP process data (and NAEP item maps with information about item difficulty) to answer questions about learning trajectories. There may be clusters of students who have different patterns of performance depending on the item difficulty that could highlight alternative learning trajectories. The data could be used to test commonly held hypotheses such as the need for demonstrating knowledge of fractions before succeeding in algebra and whether SWD are exhibiting the same patterns as students without disabilities.

Melina Uncapher of the University of California San Francisco described how researchers could use NAEP process data to understand, categorize, and model different root causes in math challenges. Neuroscience data tell us that the brain networks involved with math learning overlap with those for other cognitive tasks, including language processing and executive functioning, and thus NAEP process data may be particularly useful for examining important relationships between executive functioning, language processing, and math ability. For instance, two students who show the same performance on a math task may be exhibiting that performance for different reasons. One may struggle with language processing, whereas the other may struggle with identifying the important parts of the task and ignoring the task-irrelevant parts. By using NAEP process data to begin to understand mechanistic relationships between mathematical learning, language processing, and executive functions, we can begin to build phenotypes and patterns of relationships across student approaches to mathematical thinking and across types of disabilities. Dr. Uncapher gave the example that some studies have shown that a region in the left parietal lobe is involved in retrieval of known math facts, while arithmetic procedural processing is supported by a more distributed parietal network. To the extent that studies using the NAEP process data could untangle the retrieval of math facts from procedural fluency, for instance, this knowledge could inform the design of interventions that independently target each process based on individual student needs. While neuroscience could inform and guide research using the NAEP process data, the hypotheses that are generated and resultant interventions would need to be tested in future studies. This type of multifaced approach would be very powerful.

Discussion

A question was posed about whether there is overlap between the NAEP and what is being taught in eighth grade math, in terms of the types of questions and response demands. Dr. Davenport commented that prior knowledge affects how students process and respond to questions. People with prior knowledge become efficient at interpreting information and knowing how to respond. In contrast, a student who has not has not been exposed to certain question or response types (such as drag and drop) in mathematics class might have trouble using this on the NAEP assessment. The difficulty is that we don’t know exactly how these students are instructed. Dr. Uncapher added that it is important to differentiate between whether students are understanding the prompt and whether they are carrying out the mathematical operation we would expect. This is an opportunity to use insights from cognitive science to engage with teachers about different bottlenecks in the processing and expression of a skill. Researchers could pull out some of the larger patterns of problems from the NAEP data to use when discussing with teachers how to teach math to improve outcomes for SWD. This work could inform professional development for math teachers.

Data Mining Using Process Data from Digital Assessments

Sidney D'Mello of University of Colorado, Boulder offered the perspective of a computer scientist on how to approach research questions using the NAEP process data. He described some broad categories of research questions, including counts and correlations, comparison, prediction, and clustering. For example, comparisons can be useful to find the smallest set of generalizable behaviors (such as sequences of clicks, UDL features used) that explain differences between groups of interest. He also explained the different analytic techniques that could be used for these types of research questions. For example, predictive modeling could be used to determine whether click-stream data predicts outcomes of interest. Sequence modeling and methods to discover multivariate associations and dependencies could also be useful with this data. However, it is also important to note that these behaviors only provide one lens into internal (in the mind) processes and may not be as successful in capturing how they feel, what they think, and their motivation when they are taking the assessment that are not captured in the data.

Dr. D'Mello also cautioned that the risk of finding something obvious is very high, which is why teams that use this data should include researchers with deep content knowledge and specific hypotheses working alongside researchers who understand the data science methods needed to test those hypotheses. Ultimately, pairing people with good content knowledge who know which questions to ask with people who know the latest methodologies will maximize the usefulness of the NAEP data.

Andy Krumm of the University of Michigan, Ann Arbor shared insights from experience working with large datasets and developing measures of learning behaviors using process data from digital learning environments. He suggested that evidence-centered design (ECD) may serve as a useful framework for making sense of process data that we have about what students are doing as they complete the NAEP items. According to the ECD framework, there are constructs described as primary drivers (such as mathematics ability) and additional drivers (such as ways of thinking) as well as task features (such as a drag and drop response type task) which influence what is observed (such as a correct or incorrect response to an assessment item). This same logic can be applied to claims that are made using process data. Dr. Krumm cautioned that naming a construct once it has been quantified is moving into the world of measurement and that there are methods for strengthening one's measurement claims.

Discussion

A question was raised about the best approach for naming constructs or patterns observed in the response data. Dr. Krumm recommended being as straightforward as possible and operationalizing the construct by bringing meaning to what is being quantified rather than being poetic or alliterative. He also recommended running the explanation through the ECD process since you might think something is driving the observation until you have to go through the process of putting together an entire evidence claim related to it. Dr. D'Mello also recommended considering it from the perspective of a psychometrician who says if this construct is this, how should it predict that? However, what would work best would be to extrapolate to data outside of the NAEP assessment and then use that information to adjust the assessment.

Overview of the NAEP Process Data

Ruhan Circi of American Institutes for Research provided an overview of the NAEP process data. She also discussed the challenges of pulling out meaning, especially on SWD, from this type of data. Other topics in this presentation included how the process data were prepared for analysis and examples of analyses that have used NAEP process data. This information is not included in this summary because the full presentation will be provided as an on-demand webinar on the IES Research Funding Opportunities On-Demand Webinars website (https://ies.ed.gov/funding/webinars/index.asp).

Eighth Grade Mathematics and Students with Disabilities

Russell Gersten of Instructional Research Group provided some background on how SWD receive math instruction. He explained that the goal of the Individuals with Disabilities Education Act is to offer meaningful access to the general curriculum for students. This goal is challenging for many teachers of grades 6 to 8. Some students with IEPs in mathematics are taught in general education classes, while others are taught in special education settings, such as resource rooms. In the co-teaching model, a mathematics teacher and special education teacher work together; often they do not have common planning time or a coherent curriculum. Much of the material from grade-level textbooks is too difficult and the lessons move too quickly for students to absorb subtler ideas. In the other model, resource rooms, SWD are pulled out of their classes for individualized mathematics instructions where they often receive instruction by a computer. Both models have limitations. Ultimately, access to the grade-level general education math curriculum is difficult for SWD because they often struggle with foundational math concepts. The NAEP survey and process data can help researchers understand the types of assessment items or areas of mathematics that are most challenging for eighth grade SWD. The data could also help identify areas of strength for both students and teachers. This information could guide the development of new interventions.

Anne Foegen of Iowa State University reported that SWD often have trouble representing mathematics information in a problem and processing that information to reach a solution. By eighth grade, the gap between SWD and their peers widens as content becomes more complex, especially in algebra. Even students with no known disabilities struggle with fractions, integer computations, and fluency with number combinations. Also, how much teachers focus on procedural skills, conceptual understanding, problem solving, or reasoning varies widely. Some teachers underestimate the abilities of SWD; however, these students sometimes outperform their peers on conceptual assessments because their disability impairs only their procedural skills and knowledge. Dr. Foegen recommends researchers examine how students' familiarity with the digital administration context impacts their test-taking behaviors, including use of tools, as well as their performance; the NAEP questionnaire and process data could be used to support these investigations. For future research and funding opportunities, it would be good to consider the use of cognitive lab studies and interviews using think-aloud protocols with students with disabilities to understand how they work through the NAEP interface and their thinking processes and to confirm any hypotheses that come out of the work with the process data.

Rajiv Satsangi of George Mason University explained that SWD often make many procedural errors. Many of these students do not understand theoretical ideas related to procedure use and they have trouble doing different steps in the right order for any type of problem. Students who tend to struggle the most with the eighth grade curriculum are those with difficulty counting, identifying numerals, comparing magnitude, and working memory. For instance, SWD often have trouble recalling basic arithmetic facts, which is a particular problem for algebra. They often make retrieval errors and have varied reaction times when they retrieve facts. These students often struggle with spatial representations of numbers and might misinterpret visually represented information. However, these limitations can be overcome through instruction. These students often learn best with explicit instruction on the steps they need to take, several different representations of the same concepts, cognitive strategies (such as schematic diagrams to organize information in word problems), and metacognitive strategies (such as self-monitoring checklists to recall the steps in a problem).

Dr. Satsangi proposed several research ideas involving the NAEP process data. Researchers could examine differences in the level of engagement of SWD when problems include several illustrations or no illustrations. There also may be differences in the features of a representation (color, size, and shape) and differences in the use of language embedded in problems that may affect student performance and engagement. Researchers could also look at how students interact with that language, such as underlining words or numbers with the highlighter tool. The findings from such analyses may provide educators with a clearer picture of how students with disabilities approach and engage with test questions and may indicate potential solutions for altering the way we assess mathematics proficiency of diverse learners.

Discussion

Many algebra teachers use virtual manipulatives or algebra tiles to teach abstract algebra equations. Tangibles or haptic feedback could help students who have disabilities, including those who have visual impairments, with complex mathematical operations. But research on virtual manipulatives is still at an early stage and the evidence for SWD is limited.

In most mathematics classrooms, students do not spend much time working with digital devices, so it is unclear whether there is alignment between accommodations and tools that students have access to in the classroom and ones provided in the NAEP. Research is needed to assess differences between tools used to instruct SWD and those used to measure their performance.

Ideally, analyses of the NAEP data could show whether there are systemic weaknesses in mathematics instruction, such as not enough time spent on certain topics or problem types. Research could also identify strengths of SWD and the potential value of strength-based mathematics interventions.

Common Student Misconceptions in Eighth Grade Mathematics Content

Steve Ritter of Carnegie Learning explained that some of the knowledge that students need to answer a NAEP question is non-mathematical knowledge, such as an understanding of the question's real-world context and what to do first to solve a problem. SWD might lack the flexibility or the working memory load needed to find the best approach to each question. For instance, some test questions are easier to answer for students who have memorized facts.

Researchers could also consider differences in strategies used by students, including strategies on what part of a problem to start with to help solve the problem more easily. There is a general strategy that people use which is do the easy things first and then worry about the harder things, so they get as far as they can doing the easy things, like labeling the largest part of a pie chart and then doing the harder parts later. It turns out that if you eliminate your choices enough the harder stuff ends up not being all that hard. Knowing that strategy is an important part of answering a question. It could be that there are students who are instructed in being systematic in solving problems or may have other ways of thinking about things depending on the task. The NAEP process data may not have all the data to infer the kinds of strategies being used, but it does include processes like timing, order of steps, and errors, which could be used to understand some of the strategies and then to consider how these strategies differ by student characteristics. Researchers could explore whether the accommodations being provided support or frustrate the application of certain strategies.

Recommendations for Conducting Research using NAEP Data

During a lightning round, participants encouraged IES to consider the following for a grant competition focused on the NAEP process data:

- Provide training for researchers that includes overviews of the NAEP assessment and

details on research that has used NAEP process data
- Require applicants to form interdisciplinary teams with the needed expertise
- Share data that has already been cleaned and processed by the contractor
- Hold a competition or hackathon to identify a rich set of derived variables that could be generated from the NAEP data
- Encourage researchers to share tools or approaches for the NAEP data
- Hold regular meetings for those working with the data to share insights and lessons learned
- Provide a mechanism for the researchers using this data to provide feedback to NCES to inform future iterations of the assessment
- Provide small grant opportunities for postdoctoral fellows or graduate students to use the NAEP data

Participants also proposed the following research priorities and considerations for the NAEP data:

- Consider the context, including student characteristics and the cognitive processes required to respond to a question
- Focus on research that will inform classroom instruction
- Examine relationships between variables in student and teacher surveys and item performance by SWD (combined with focus groups or interviews to confirm the findings)
- Focus on research that would allow for a better understanding of assessment practices, including how accommodations and UDL features maintain validity of the test and contribute to improved performance
- Analyze data on individual NAEP assessment items and be cautious with generalizations across all items
- Attend to the alignment between classroom instruction and assessment and the digital NAEP

Appendix. Agenda and Expert Briefs

Prior to the Technical Working Group Meeting on March 18, 2020 the invited experts provided short written briefs. The purpose of these brief was to guide the meeting agenda and to provide IES with recommendations for a NAEP process data competition. The agenda for the meeting and these expert briefs are included in this appendix.

NAEP Mathematics Data for Students with Disabilities
Technical Working Group Meeting
Institute of Education Sciences
March 18, 2020

AGENDA

Objective: IES is working to release process data (including keystrokes and use of tools and accommodations within the assessment) from the 2017 eighth grade mathematics NAEP in addition to other survey data collected from learners, teachers, and schools. The purpose of this Technical Working Group meeting is to identify areas of research that are needed and could be addressed with this data as well as data science approaches that could be used in such research.

9:00-9:05 **Virtual Meeting Directions and Tips for Meeting Success**

- **Sarah Brasiel**, Program Officer, NCSER

9:05 - 9:20 **Welcome**

Presenters:

- **Mark Schneider,** Director of the Institute of Education Sciences (IES)
- **Joan McLaughlin**, Commissioner, National Center for Special Education Research (NCSER)
- **Lynn Woodworth,** Commissioner, National Center for Education Statistics (NCES)
- **Peggy Carr**, Associate Commissioner, NCES
- **Sarah Brasiel**, Program Officer, NCSER

9:20-10:20 **NAEP Assessment and Restricted-Use File(s)**

Representatives from NCES will provide a brief overview of the NAEP assessment design, the sample design, the data that will be made available to researchers, and inferences that can be made with this data.

Presenters:

- Emmanuel Sikali, NCES
- Dan McGrath, NCES

10:20-10:50 **NAEP Administration for Students with Disabilities**

In this session we will discuss the exclusion criteria for student participation in NAEP, identification of students with disabilities in need of accommodations, and some of the challenges and strategies used to administer NAEP to students with disabilities.

Presenters:

- Grady Wilburn, NCES
- Jenny Cain, Minnesota Department of Education, NAEP State Coordinator
- Lizanne DeStefano, Georgia Institute of Technology

10:50 - 10:55 Check-in on Virtual Meeting Success/Feedback to Improve the Meeting

10:55 - 11:10am **BREAK**

11:10 - 11:40 **Universal Design for Learning (UDL) and Testing Accommodations**

During this session we will discuss different features of the eNAEP platform as well as research on use of UDL and other accommodations within digital assessments used by students with disabilities.

Presenters:

- Marcelo Worsley, Northwestern University
- Leanne Ketterlin-Geller, Southern Methodist University

11:40 - 12:10 **Student Cognition and Learning in Mathematics**

Presenters in this session will discuss what is known in the fields of cognition and neuroscience that can inform research using the eighth grade NAEP released items and related process data.

Presenters:

- Melina Uncapher, University of California San Francisco
- Jodi Davenport, WestEd

12:10 - 12:45 **Data Mining Using Process Data from Digital Assessments**

In this session we will discuss some of the types of data mining approaches that can be used to understand test-taking behavior and performance of students with disabilities in a digital environment, including the use of accommodations and universal design for learning (UDL) features of the platform.

Presenters:

- Sidney D'Mello, University of Colorado, Boulder
- Andy Krumm, University of Michigan Medical School, Ann Arbor

12:45 - 1:15 **Lunch**

1:15-2:00 **Intro to NAEP Process Data**

Process data has great potential to provide deeper insights about different elements of the NAEP assessment cycle. It can also add great value to a variety of topics relevant to special education, especially for learners with disabilities. This presentation will provide basic elements of NAEP process data with a focus on data management, quality, and security. Examples using this exciting data will be explored.

Presenter:

- Ruhan Circi, American Institutes for Research

2:00 - 2:45 **Eighth Grade Mathematics and Students with Disabilities**

In this session we will discuss what is known about challenges with this level of mathematics for students with disabilities and what is needed to understand testing behavior and performance of students with disabilities on different item types.

Presenters:

- Russell Gersten, Instructional Research Group
- Anne Foegen, Iowa State University
- Rajiv Satsangi, George Mason University

2:45-3:00 **Break**

3:00 -3:15 **Common Student Misconceptions in Eighth Grade Mathematics Content**

This session will highlight some of the challenges that students with and without disabilities face at this level of mathematics and some of the common student misconceptions related to the eighth grade released NAEP items.

Presenter:

- Steve Ritter, Carnegie Learning

3:15- 3:55 **Recommendations for Conducting Research using Data from NAEP**

This will be a lighting round session where TWG members will be asked to share their highest priority for guiding future funding opportunities utilizing NAEP process data for students with disabilities.

Group Discussion

3:55 - 4:00 **Closing Remarks**

Presenter:

- Joan McLaughlin, Commissioner, National Center for Special Education Research

Summary of Issues Related to NAEP Administration for Students with Disabilities

Jennifer Cain, NAEP State Coordinator
Minnesota Department of Education

Before delving into any analysis of NAEP data, it is important to understand how NAEP works, more specifically how the assessment is administered. As a NAEP State Coordinator (NSC), I assist selected schools through the NAEP administration process, from notification to administration. This brief summary provides details about the pre-administration process for students with disabilities.

Administration Logistics. NAEP policy states that all student information must be verified at the school level. This verification is completed by a school staff member who is designated the school coordinator (SC) for NAEP during the pre-assessment activities. There are approximately seven pre-assessment activities the SC must complete to prepare for the administration at the school, including identifying students with disabilities and identifying necessary accommodations. The SC uses the MyNAEP website to complete all pre-assessment activities as well as phone conversations with their NAEP representative, a member of the NAEP Field Staff who will be administering the assessment at the school.
The first task of the pre-administration process is for each school to submit a list of students. This list must contain all enrolled students within the selected grade at the school. The NAEP contractor uses this list to randomly select the student sample, usually a sample size of 50 students. Each student on the list has an equal opportunity of selection, regardless of disability status.

Identifying Students with Disabilities. After the list of students selected for NAEP is loaded onto the MyNAEP website, the SC must verify that all student information is correct, this includes identification as a student with disabilities. The SC is given the NAEP definitions for students with disabilities to ensure that each selected student is identified correctly according to the NAEP definition (See Supplemental NAEP Definitions).

Identifying Necessary Accommodations. After a student is identified as a student with disabilities, the next step is for the SC to identify the accommodations needed by the student for taking the assessment. The first step of this process is for the SC to review the State's NAEP Inclusion Policies.

State's NAEP Inclusion Policies. Since NAEP is administered universally across the United States, the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES) create Inclusion Policies for each NAEP assessment. The Inclusion Policies include the eligibility criteria for students with disabilities (SD) and the Universal Design Elements (UDEs) and testing accommodations available for the NAEP assessment.

However, to ensure that students can access the assessment, the NSC edits the inclusion policies for their state. NSCs edit the policies by removing any accommodations that are not allowed on the state assessment. The accommodations are removed because if the state assessment does not allow a specific accommodation than a student would not be accustomed to testing using that accommodation; and having a student use an unfamiliar accommodation could affect the results. The NSC does not remove any of the UDEs because the student tutorial, which the student receives at the beginning of the assessment, provides instructions to the student on how to use the UDEs, so the student is familiar with them, regardless if the UDEs are available on the state assessment.

> *State Context:* For the NAEP 2017 Mathematics administration, the Minnesota Inclusion Policies did not include the "Cueing to Stay on Task" accommodation. This accommodation was removed because the description of the accommodation is not allowed on the Minnesota Comprehensive Assessments (Minnesota's State Assessment). Therefore, the accommodation was not allowed for any students selected to take NAEP in Minnesota.

Assigning Accommodations. After reviewing the State's NAEP Inclusion Policies, the SC will assign the UDEs and accommodations needed by the student to take NAEP. The SC is asked a series of questions to determine the best way to assess the student on the NAEP assessment (see NAEP Students with Disabilities Questionnaire). The student should be assessed on NAEP in a similar way that they are assessed on the State Assessment. If the student should be assessed with accommodations, the SC will select the UDEs and accommodations the student needs to take NAEP.

If the student needs an accommodation that is not included in either the UDEs list or NAEP accommodations list, the SC may specify the accommodation. The requests are reviewed by the NAEP contractor and the SC is notified whether the accommodation is allowed or not allowed by NAEP. If the accommodation is allowed by NAEP then the student will be assessed on NAEP with the accommodation. However, if the accommodation is not allowed by NAEP, a decision will have to be made on how the student should be assessed by NAEP, whether it be assessed without the necessary accommodation or excluded from NAEP.

Student Exclusions. Part of the pre-administration process is to decide how the student should be assessed on NAEP, meaning should the student be included or be excluded from NAEP. The decision is made by the parents, student, and/or school; and it can lead to student exclusions. There are two classifications of student exclusions: excluded in accordance with NAEP Inclusion Policy (see Appendix A) and excluded contrary to NAEP Inclusion Policy.

The NSC monitors all exclusions with the goal of ensuring all students selected take NAEP and minimizing the number of students that are excluded contrary to NAEP Inclusion Policy. The

NSC will contact the SC if any student is excluded contrary to NAEP Inclusion Policy. The NSC encourages all selected students to take NAEP; however, it is understood that there may be special circumstances and each exclusion is considered on a case-by-case basis with special thought given to the student.

> *State Context:* In Minnesota, the state assessment is fully digital, but an allowed accommodation is a paper version of the test. There have been several cases where a selected student for NAEP requires this accommodation; but a paper version of the test is not a NAEP available accommodation. In these cases, I discuss with the SC and the Special Education Specialist that works directly with the student to determine if the student should take NAEP. We consider the student's ability to access the assessment as well as the impact of taking NAEP on the student's mental, behavioral, and physical health. Although, I encourage that all students participate in NAEP; I can support a decision of exclusion if the case mandates it.

Conclusion. This summary provides the process that selected schools complete to identify students with disabilities and identify any necessary accommodations. The state context was provided so researchers are aware that there are small differences in NAEP administration for students with disabilities across the nation, which is important to be mindful of when analyzing NAEP data at the national level.

Supplemental NAEP Definitions for Students with Disabilities

NAEP defines a student with disability as a student with an Individualized Education Program (IEP) for reasons other than being gifted and talented or a student on a Section 504 Plan that requires accommodations to access the NAEP assessment. The categories are

- **Yes, IEP.** Student has a formal Individualized Education Program (IEP), the student's IEP is in progress, or the student has an equivalent classification for private schools. If a student has both an IEP and a 504 Plan, code these students as "Yes, IEP".
- **Yes, 504.** Student has a 504 Plan and requires an accommodation for NAEP. If a student has both an IEP and a 504 Plan, code these students as "Yes, IEP".
- **No, not SD.** Student does not have an IEP and/or a student has a 504 Plan and does not require a testing accommodation.
- **Information unavailable at this time.** Information is currently unavailable. Please note that your NAEP representative will need to collect the data at a later date, if possible.

Student Exclusion Criteria from the NAEP Inclusion Policy

Students identified as students with disabilities can be excluded from NAEP if the student meets (or met) the participation criteria for the State Alternate Assessment(s) (i.e., the State's alternate assessment based on alternate achievement standards).

NAEP Students with Disabilities Questionnaire

Complete Students with Disabilities Student Information
To ensure that NAEP reflects the educational progress of all students, students with disabilities (SD) must be included to the fullest extent possible.

NAEP provides accommodations only for SD. Therefore, this section only includes students whose SD status is "Yes, IEP" or "Yes, 504."

The information that you provide in this section indicates how students will participate in NAEP and the accommodations they will receive.

Review NAEP Inclusion Policy
NAEP expects most students with disabilities to participate in NAEP. The only students eligible for exclusion are students who meet (or met) participation criteria for the state Alternate Assessment. All other SD students should participate with or without NAEP accommodations.

Your NAEP State Coordinator will monitor exclusions and may contact you regarding excluded students. If you have any questions about how to include a student, contact your NAEP State Coordinator.

Provide Information for Students with Disabilities

Students with Disabilities Details

For each student with a disability, record the following information.
Which of the following IDEA categories best describes each student's identified disability? Select all that apply.

- □ Specific learning disability
- □ Hearing impairment/deafness
- □ Speech or language impairment
- □ Intellectual disability
- □ Emotional disturbance
- □ Orthopedic impairment
- □ Traumatic brain injury
- □ Autism
- □ Developmental delay (age 9 or younger)

- □ Visual impairment/ blindness
- □ Other health impairment (Specify)

What is the degree of each student's disability(ies)?

- □ Profound/Severe
- □ Moderate
- □ Mild
- □ Don't know

At what grade level does this student perform in the NAEP subject?

- □ At or above grade level
- □ One year below grade level
- □ Two or more years below grade level
- □ Not receiving instruction in this subject
- □ Don't know

How Students Should Participate

For each student with a disability, record the following information. Review the NAEP inclusion policy prior to completing. If you have any questions on how to include students, contact your NAEP State Coordinator.

How is the student assessed on the state or classroom assessments?

- □ Without accommodations
- □ With accommodations
- □ Meets (or met) participation criteria for the state alternate assessment

How should the student be assessed on NAEP?

- □ Without accommodations
- □ With accommodations allowed or provided by NAEP
- □ Do not test

If student is assessed with accommodations allowed or provided by NAEP, what accommodations does the student need? Refer to the NAEP inclusion policy for the universal design elements and accommodations that are allowed or provided by NAEP.

Cognition-related Issues and Opportunities for Exploring NAEP Mathematics Data for Students with Disabilities

Jodi Davenport, Deputy Director for the Science, Technology, Engineering, & Mathematics
WestEd

Given the disparity in performance between students with versus without disabilities on the 8th grade NAEP exam, a fundamental question is whether the scores on the exam reflect large differences in students' math knowledge and skills or reflect challenges associated with test demands and motivation. Detailed process data from the 2017 eighth grade mathematics NAEP assessment provides the opportunity to generate new insights on what factors may be influencing performance.

Cognitive processing

Test taking requires a range of cognitive processes, e.g., attention, perception, memory, and reasoning, and the resources for these processes are limited. Students must understand what is being asked, recall and apply relevant knowledge and skills, and provide a response in the expected format. Ideally, test items minimize the need for task-irrelevant processing and students' responses reflect their understanding of the content area rather than their ability to navigate the test. Do test features create unnecessary cognitive demands that prevent students with disabilities from demonstrating their proficiency in math?

Features of assessment tasks can increase or decrease demands on attention, perception and working memory, and these demands may disproportionately affect students with disabilities. Attentional and perceptual demands are increased when displays are crowded, irrelevant representations are included, or multiple prompts appear on a single page. Working memory is required for students to hold in mind the relevant pieces of information to complete a task. Demand increases if students must integrate information from multiple representations, e.g., text and a diagram, or are limited in their ability to off-load information while solving a problem, e.g., by writing down intermediate steps on paper. Accommodations, such as converting visual representations to text-to-speech may increase processing demands for students with disabilities. Other test features may reduce processing demands, such as the ability to use a highlighter, or adjust the visual contrast of the item.

Once students understand the task and have an answer, they must provide a response in the format expected by the test developer. Students with disabilities may find it more challenging to provide the intended response due to interface constraints. If multiple prompts are provided on the same page, attentional resources are required to select the correct place to put the corresponding answer. In other cases, the response format itself may introduce difficulties. A drag-and-drop response may be intuitive for sighted students, but students using a screen reader may struggle to hold options in working memory. Other features of the NAEP exam, such as the ability to eliminate options from multiple choice likely reduce processing demands.

Opportunities

The detailed process data available from the 2017 NAEP exam along with corresponding survey and demographic data allows researchers to explore questions related to the influence of processing demands on performance, test-taking motivation, and math learning trajectories. Do students with disabilities struggle with particular features of items or response types? Process data could explore whether features, such as multiple prompts on the same page, or response types, such as drag and drop, affect performance for students with different types of disabilities. Recommendations could follow, such as, presenting test items sequentially rather than all on one screen.

Are accommodations helping students with disabilities? Process data could show how the use of tools, such as removing multiple choice options, text to speech, or changing visual contrast, relates to time on task and overall performance.

Are students with disabilities actively engaged in test tasking? Cognitive demands are only relevant if students attempt to respond to the questions. If students are not motivated, they may exhibit behaviors such as guessing or skipping answers. Process data related to the number of movements or time on each item could help tease apart students that are struggling from students that may have motivational issues. The detailed data enable researches to look at test taking patterns over time that may reveal initial engagement that later morphs to fatigue or distraction.

Do different subsets of students follow similar learning trajectories? Research in math education suggests that certain concepts, e.g., an understanding of fractions, are important precursors for later math learning, e.g., algebraic problem solving. The released NAEP Item Map orders items reflecting math concepts by difficulty. Process data could reveal whether these concepts progress similarly for all students.

Overall, the availability of process data may enable researchers to deepen our understanding of features affecting mathematics performance, inform the design of future assessments, and provide more equitable opportunities for students with disabilities to demonstrate their knowledge and skills.

Exclusion Criteria: Identification of SWD in need of accommodations

Lizanne DeStefano, Professor of Psychology and the Executive Director of the Center for Education Integrating Science, Mathematics, & Computing-Georgia Institute of Technology

Critical issues we should consider:

- Are the Student Questionnaire and the Teacher Questionnaire **appropriate** for use with SWD and special education teachers? Are there additional questions that should be asked in future administrations to better understand differences in curriculum, pedagogy and educational environment across regular and special education requirements?
- It would be very useful to get a good **description** of the group of SWD who do and DO NOT take NAEP. A careful analysis of data from the SWD worksheet and student questionnaire in conjunction with performance data might shed some light into how decisions are made to include/exclude students from NAEP and, if included, which accommodations are assigned/used by the students.
- **Comparisons** of student, teacher and school characteristics between students who are excluded from NAEP and those that are included would allow us to understand the exclusion decision-making process, its consistency from school to school and implications of the validity of NAEP for SWD.
- **Analysis** of patterns of accommodations assigned (SWD worksheet) vs. accommodations used (process data) would give some insight into how students actually engage during the testing situation.
- **Understanding any relationships** between accommodations used, test taking behavior, and item type would contribute to our understanding of how students approach the test. Linking that with success or failure on the item could inform our understanding of value-added of accommodation.
- **Comparison** of math instruction (from student and teacher questionnaires) between SWD (included and excluded) and non-SWD students and then association of differences in instructional variables with performance on specific items (or by performance level) would illuminate key differences in the math preparation of SWDs vs. grade level peers.
- **Comparison** of test taking behavior between SWD and non-SWD and correlation of test taking behavior with instructional characteristics and performance on specific items would help us understand the unique performance patterns of SWD.

Opportunities to help us guide researchers in conducting analyses using the NAEP data:

- It is always difficult to study special populations performance on NAEP because of small sample sizes and the matrixed nature of the test. **Published guidance** on how SWDs are distributed across blocks in NAEP and the content of those blocks would assist researchers in selecting blocks for study.
- Perhaps a **white paper** on how SWDS are represented in NAEP and a portal that brings together relevant data sets for research on SWD would be a good start.
- A **webinar or boot camp** on how to conduct analysis of NAEP data for SWD might be a nice event to promote use.
- **Small grant competitions** to support use of NAEP data to study SWDs could engage some graduate or post doc researchers. These could be linked to the boot camps mentioned above.

Data Mining Using Process Data from Digital Assessments

Sidney D'Mello, Associate Professor
University of Colorado, Boulder

How can NAEP process data be used to understand test-taking behavior and performance of students with disabilities? Below are potential analyses, each focused on a particular research question, followed by an overview of analytic techniques. A predictive modeling approach is applied here where the goal is to produce a computational model that predicts an outcome of interest (e.g., test score) from click-stream data (behaviors extracted from NAEP log-files). Critically, the model needs to be generalizable in that its predictions are accurate when tested on out-of-sample (held out) data. This is the key distinction from traditional analytical methods which focus on explanation rather than prediction (Yarkoni & Westfall, 2017).

Analyses

1. **Count and Correlate**. The most basic approach would be to count each behavior (e.g., "Horizontal Item Scroll") or sequence of behaviors ("Horizontal Item Scroll" ➔ "Highlight") and correlate them with pertinent variables of interest, such as disability accommodations, performance on specific items, etc. In addition to working with the raw behaviors, higher-order latent representations can be automatically extracted and analyzed.
2. **Comparison**. The idea is to identify a minimal subset of behaviors that explains differences between groups of interest. This might involve, for example, comparing patterns of behavior of students with and without disabilities or comparing students with different types of accommodations to infer differences in how each group approaches a problem and/or performs on the problem. This will require establishing a sampling frame by matching students in terms of sociodemographics, location, and other covariates; but allowing focal variables (e.g., disability or not, accommodation type) to vary. The analysis will focus on quantifying differences among groups and understanding which behaviors best explain the group differences.
3. **Prediction**. The idea is to develop an accurate and generalizable predictive model of an outcome of interest, for example, performance on an individual item or performance on the entire assessment as a whole. The extent to which click-stream data can yield accurate predictions provides an index into its information value. Importantly, the model should be constructed so as to generalize across students and student groups of interest. An analysis of the model parameters and structure can provide insights into the underlying behaviors (e.g., Bartlett, Littlewort, Frank, & Lee, 2014).
4. **Clustering**. The goal is to group students or group actions within a student based on their behaviors and then attempt to understand (or label) the different groups. Clustering can be done on static data (counts of behaviors across a time window) or on dynamic data (clustering of trajectories of behaviors). This discovery-oriented technique can help uncover any latent (or hidden) subgroups in the data.

Analytical Techniques
Below is an overview of some of the techniques that can be applied to the NAEP data for the analyses suggested above. See D'Mello (2020) and Baker and Yacef (2009) for more details.

Relationship Discovery. Click-stream data is transactional - i.e., multiple "events" occur within a time frame (transaction). Much can be learned by identifying "interesting" associations from these transactions. It is particularly informative to identify multivariate associations (e.g., actions A, B, C, and D have a certain likelihood of co-occurring), which is different from multivariate correlational analyses, where two sets of variables are correlated. It is also important to go beyond associations by identifying dependencies (i.e., A implies B, but B does not imply A). Because the data can contain millions of transactions, methods to discover associations and dependencies should be computationally efficient. The field of association rule mining (Agrawal & Srikant, 1994) provides methods to identify multivariate associations and their dependencies (e.g., [A, B, C] ➔ [D, E]; events A, B, and C co-occur and predict events D and E).

Sequence Modeling. Click-stream data is also sequential. Techniques to analyze sequential data can be grouped into four categories. The first uses dynamical probabilistic graphical models, such as dynamic Bayesian nets (e.g., Conati & Maclaren, 2009), Hidden Markov Models (HMMs) (e.g., Boyer, et al., 2009), conditional random fields, and hierarchical variants of these models. The second approach is to use recurrent neural networks (RNNs) for sequence learning. RNNs are variants of feed-forward neural networks with recurrent links (e.g., backward links between output neurons and hidden neurons). The third approach is to find frequently occurring subsequences among a set of sequences (e.g., A ➔ B is a subsequence of A ➔ B➔ C and also of A ➔ J➔ K ➔ K ➔ C ➔ B ➔ L) using fast algorithms designed for this task (Srikant & Agrawal, 1996). The fourth approach is to utilize techniques from time series analyses, such as lag-sequential analysis, recurrence quantification analysis, growth curve modeling, dynamic time warping, etc., to identify sequences in the data.

Predictive Modeling. The basic idea of this type of modeling is to utilize features from click-streams to predict outcomes of interest. The features can include counts of particular behaviors, frequently occurring behavioral clusters, or even sequences of behaviors (e.g., Hutt, Grafsgaard, & D'Mello, 2019). Standard supervised learning methods, such as linear/logistic regression, nearest neighbors, decision trees, Naïve Bayes classifiers, support vector machines, feed-forward neural networks, and ensemble methods like Random Forest, can be used for the predictive modeling task, but this requires supervision in the form of labeled data (e.g., disability or not).

Unsupervised Learning. Large-scale datasets are usually heterogeneous, so some grouping based on student characteristics, contextual factors, or action sequences is warranted. There are also situations when there is an interest in identifying subgroups with certain characteristics. This is done using unsupervised learning methods -mainly cluster analysis- using standard techniques such as k-means clustering, expectation-maximization, and self-organizing maps (see Vellido, Castro, and Nebot (2010) for a brief review of clustering applied to educational data). Feature extraction and representation learning methods are also unsupervised learning techniques that can be used to extract generate higher-order latent representations in the data (Goodfellow, Bengio, & Courville, 2016).

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke & C. Zaniolo (Eds.), *Proceedings of the 20th International Conference on Very Large Data bases (VLDB 94)* (pp. 487-499). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3-17.

Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology, 24*(7), 738-743.

Boyer, K., Young, E., Wallis, M., Phillips, R., Vouk, M., & Lester, J. (2009). Discovering Tutorial Dialogue Strategies with Hidden Markov Models. In V. Dimitrova, R. Mizoguchi, B. Du Boulay & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 141 - 148). Amsterdam: IOS Press.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267-303.

D'Mello, S. K. (2020). Big Data in the Science of Learning. In S. E. Woo, L. Tay & R. Proctor (Eds.), *Big Data methods for Psychological Research: New horizons and challenges*. Washington, DC: American Psychological Association.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT press.

Hutt, S., Grafsgaard, J., & D'Mello, S. K. (2019). Time to Scale: Generalizable Affect Detection for Tens of Thousands of Students across An Entire Schoolyear *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2019)*. New York: ACM.

Srikant, R., & Agrawal, R. (1996). *Mining Sequential Patterns: Generalizations and Performance Improvements.* Paper presented at the Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology.

Vellido, A., Castro, F., & Nebot, A. (2010). Clustering educational data. In C. Romero, S. Ventura, S. R. Viola, M. Pechenizkiy & R. S. J. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 75-92). Boca Raton, FL: Chapman & Hall/CRC Press.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122.

Eighth Grade Mathematics and Students with Disabilities

Anne Foegen, Professor and Director of Graduate Education in the School of Education
Iowa State University

I appreciate the opportunity to serve as a member of this technical working group. To provide context for my analysis, I begin by sharing a bit about my background and research. As an undergraduate, I majored in secondary mathematics teaching, with minors in learning disabilities and mild/moderate intellectual disabilities. My master's work was in gifted education, and my doctoral work in special education, primarily learning disabilities. My research has involved the development and implementation of progress monitoring tools in mathematics at the early (K-1) grades, middle school, and high school. I have had the great fortune to work closely with colleagues in mathematics (general) education in research and professional development contexts. Over the past 15+ years, I have been focused on algebra (early algebra and Algebra 1), and have worked in with middle and high school teachers across the country. These experiences with general and special education mathematics teachers in professional development settings and in their classrooms, along with my colleagues in special and general education, shape my responses. In the comments below, I use SWD to refer to students with disabilities; given my primary focus on academic disabilities, it is likely that I am not attending to issues related to other types of disabilities.

What is known about challenges with eighth grade mathematics for students with disabilities?

The challenges that students with learning/academic disabilities experience in eighth grade mathematics are often associated with difficulty with representing mathematical information (e.g., interpreting what a problem requires) as well as processing information to arrive at an answer. Research studies within cognitive psychology have identified retrieval and long-term memory representations for basic facts, delayed acquisition of procedures for computation, and number representation (e.g., on a number line) and processing (e.g., comparing number magnitudes) as common areas of difficulty. These difficulties with mathematics learning through the elementary years often result in SWDs reaching grade eight with significant gaps in the prerequisite skills needed for success in mathematics just as the curriculum is making greater demands for higher order thinking and problem solving.

Eighth grade students (including many without identified disabilities) demonstrate gaps in important prerequisite skills and concepts necessary for success in the mathematics content assessed by NAEP, including fractions, integers, estimation, and fluent computation. In addition, many curriculum materials used in general education may assume all students are fluent with computation and retrieval of basic facts. For students with other types of disabilities (and those with LD), attention, persistence with a task, and motivation may play a significant role in their success in engaging in eighth grade mathematics.

The nature of mathematics instruction (both curriculum materials and instructional practices) is also an important consideration. Scholars in mathematics education and special education often value different aspects of mathematics learning and performance. Mathematics educators are more likely to emphasize conceptual understanding, reasoning, and communication of mathematical thinking, while special educators often focus more on procedural fluency and solution accuracy as they deliver specially designed instruction to SWDs. Despite considerable effort among mathematics education scholars, teacher educators, and professional organizations, my experiences in schools across the country have included very few instances in which the classroom instruction in general education mathematics aligns with the goals and values of the mathematics education community. The experiences of eighth grade SWDs in mathematics will be influenced by the nature of the instruction they have received and, particularly if they do not receive mathematics instruction in general education settings, it may not align with the grade level standards and therefore, the NAEP Mathematics Framework.

With regard to curricular and instructional demands for conceptual and procedural learning, I have noticed that teachers (both special and general education) often underestimate the SWDs' abilities to engage successfully with conceptually-oriented tasks and assessments. In my experience, some SWDs demonstrate greater levels of success with contextualized problems (which teachers often judge as having "too much reading") and conceptual tasks (which place less emphasis on procedures and algorithms that they struggle to retrieve.

What is needed to understand testing behavior and performance of students with disabilities on different item types?

After reviewing information about the NAEP framework, items, and digital delivery, I see several areas for future research, though some of these extend beyond the keystroke/process data and others likely pertain to students without identified disabilities as well as SWD. Several of my questions address the degree to which familiarity with the digital administration context impacts testing behavior and performance. Data from the survey items about familiarity with using digital tools (e.g., a tablet, a scientific calculator, an equation editor) could be used in conjunction with process data related to response times and use of the tools to explore the degree to and ways in which students with varying levels of familiarity make use of the tools.

Many of the questions I have likely require data that isn't available at the student level. As examples: How do the types of universal design aspects (e.g., read aloud, enlarge text, highlight) the student accesses on NAEP compare to those provided in the student's IEP accommodations? Does the type of calculator a student regularly uses in the classroom impact the ways in which the NAEP online scientific calculator is used? Does the process data include the scratch work that was done, or just when the tool was turned on and off? The ability to view scratch work would produce a rich set of data for analyses of SWD's problem solving approaches and possibly misconceptions. Have any process studies been done to understand how SWDs are interpreting the items and pursuing solution paths? The ability to conduct such studies and then use them to interpret response data from NAEP could provide valuable insights for intervention development. Is there any IEP data that is more fine-grained, or is this a dichotomous variable? It would be valuable to distinguish SWD with different types of IEP goals (e.g., mathematics, reading, written expression, behavior) when examining performance and testing behaviors.

Middle School Mathematics Instruction for Students with Disabilities

Russell Gersten, Executive Director
Instructional Research Group

This perspective is shaped primarily by my experience working with two middle schools in two different regions on a model demonstration project involving middle school mathematics instruction for students with disabilities and other low performing students. Although sample size is small, the issues seem to recur in the other current projects working on this topic.

Remarks are organized around:

- Problems in service delivery options for students with disabilities for mathematics
- Recurrent observed problems that students experience and teachers' reactions.
- Implications from other relevant research

Problems in service delivery options for students with disabilities for mathematics

The three prevalent service delivery models that schools use appear to be

a) co-teaching (with one mathematics teacher and one special education teacher teaching a class of 25 or so)
b) special education mathematics lab where students with IEPs work on programs such as Istation that allow for individualized instruction and frequent feedback. Teachers typically supervise and perhaps provide assistance if students appear to need help.
c) Schools may also use the model of inclusion classrooms using explicit math instruction delivered by a special education teacher for those students whose skill level is too low to allow their enrollment in a co-taught, general education class. In this case, the math content may remedial/foundational

These models are understandable.

Co-teaching, in particular, appears to directly meet the challenge of IDEA to provide meaningful access to grade level curriculum. By having a second teacher in the room, students can be provided with additional assistance and also potentially work on foundational skills and concepts that underlie grade level content (e.g. review of fractions if topic is proportional reasoning). One major problem is the wide range of students' mathematics proficiency in one of these classes. For example, on STAR, range was from grade 2 to grade 6 mathematics proficiency in a typical class. A second major problem is the tendency to want to move at a decent clip to cover material for the grade level. As the research on successful interventions shows, for intricate topics such as fractions, base ten, decimal fractions, word problems, students with disabilities require a good deal more practice and more feedback and explanation and review than typical students. Another problem frequently noted is that in some cases, the special education teacher essentially serves as an aide. That was not a problem in our site. Time for planning is another major problem, one that we have worked with teachers and principals on.

But major problems persist. *These include inability to devise strategies to help students with difficult topic areas such as: word problems, fractions and understanding of key mathematical ideas involved in fraction arithmetic, even seriously limited understanding of the meaning of the equal sign and virtually no understanding of the understanding of the meaning and use of the term variable.. The latter, of course, is crucial for solving equations involving more than one step, sets of linear equations, and of course, the development of algebraic reasoning.* Attempts to address these areas are often fragmented, improvisatory, and frustrating for both teachers and students.

Another recurrent problem, and one likely to be open to study on NAEP performance data is lack of task persistence/self-regulation, motivation, learned helplessness. These are profound issues. In a mainstream mathematics class, students with disabilities are often graded based on effort, an elusive concept. This concept backfires when students actually take state assessments or end-of-course exams or participate in NAEP, because standards are based on actual performance.

Problems with the computerized approach would seem fairly self-evident for a group of students who often get easily frustrated and show limited interest and motivation in mathematics. Although some schools may use some type of blended or hybrid learning approach, we have not witnessed this and see this as potentially challenging.

In terms of topics to explore in depth in NAEP:

1. Because proficiency with fractions tend to be a strong predictor of success in Algebra 1 and beyond, it might be useful to create a composite score on this aspect of NAEP. And explore what it correlates with.
2. As previously mentioned, use of NAEP keystroke data etc. to assess self-regulation/impulsivity and task persistence.
3. Examining any strong points for this group on 8th grade NAEP. Given the motivational problems that students and often teachers experience, would be good to know if there are any areas of strength. Of course, this information could be used to let teachers know which areas are not in need of extensive intervention and instruction, which can help guide teachers to be able to spend time on the complex, challenging topics.
4. Possible correlations between NAEP scores and 8th grade special education delivery model (e.g. computer-assisted instruction, mainstream core class, “basic mathematics” class, possibly with or without coteaching.)

Accommodations and Universal Design for Assessment

Leanne Ketterlin-Geller, Professor and Director of Research in Mathematics Education
Southern Methodist University

Key Tenets for Accommodations and Universal Design for Assessment

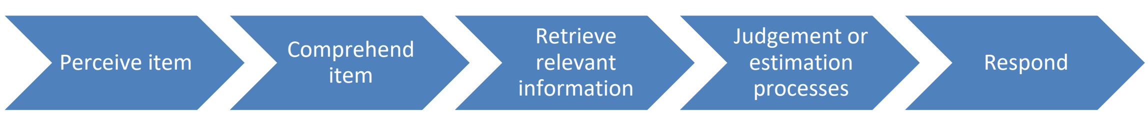
Accessibility of educational assessments refers to students' ability to engage with the tested content in a way that allows them to accurately demonstrate their knowledge, skills, and abilities (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), and is a necessary condition for validity. Inaccessible assessments lead to inaccurate interpretations, and manifests as construct irrelevant variance in students' scores, and may be difficult to detect.

Accessibility of educational assessments is impacted by the interaction between two key components inherent in every testing situation: the test taker's personal characteristics and item or test design features. Personal characteristics are typically classified into four categories: cognitive processing, attention, language or linguistic processing, and physical characteristics (Ketterlin-Geller, Crawford, & Huscroft-D'Angelo, 2014). Item or test design features include the ways in which items or tests are formatted for delivery, response, administration, and implementation, which are detailed in the test specifications.

Accessibility can be improved by attending to the interaction between these two components through intentional instrument development and implementation practices (Ketterlin-Geller, 2016). Integrating principles of universal design for assessment (UDA) during test development is intended to facilitate accessibility by attending to a range of user characteristics when designing the test specifications. UDA principles may be ubiquitously applied across the test or implemented on an item-by-item basis as user-selected accessibility supports. For test takers whose personal characteristics continue to negatively interact with item and test design features, accommodations can be applied. Accommodations are changes to the way in which directions and/or test items are presented, the mechanism through which students generate and/or produce a response, or the setting or timing of test administration. Properly applied, both UDA principles and accommodations maintain integrity to the tested construct.

Item Processing Model

Item processing can be represented as a sequential series of actions between the test taker and the item (Padilla & Leighton, 2017):



The test taker's personal characteristics may negatively interact with item design features at any of these phases of item processing. Unless mitigated through the application of UDA principles or accommodations, this interaction may create an inaccessible testing situation.

Historical Validation Methods of Accommodations and Universal Design for Assessment
Because the tested construct is not altered when UDA principles are applied or accommodations are implemented, scores obtained under these conditions are intended to be comparable to scores obtained by test takers under "traditional" testing conditions. Evidence is required to verify this claim and is typically obtained by examining the presence of a differential boost in observed score. This hypothesis states that when test takers for whom accessibility is compromised take two versions of the same test - one in which UDA principles or accommodations are applied and one without - their observed scores will be significantly higher on the former. Simultaneously, when test takers for whom accessibility is not compromised take the same two versions of the test, their observed scores will not be noticeably different. A statistically significant interaction effect is interpreted to indicate that (1) the application of UDA principles and/or accommodations improved the accessibility for some students, while at the same time not disadvantaging others, and (2) that both versions of the test are measuring the same underlying construct. Because accessibility is influenced by multiple factors, this blunt comparison may obscure more nuanced and subtle differences that lead to better understanding of accessibility of educational assessments.

Role of Response Process Data in Validation of Accessibility
Response process data may play an important role in examining two claims that underlie the premise of applying UDA principles and/or implementing accommodations for improving accessibility:

1. Claim about score comparability between test takers: Are the same cognitive processes being elicited throughout item processing under multiple conditions (e.g., UDA principles, accommodations)?
2. Claim about improved accessibility within an individual test taker: Does the application of UDA principles and/or accommodations mitigate the negative interaction between the test taker's personal characteristics and the item or test design features? For whom (e.g., which personal characteristics), under which circumstances (e.g., content, item, test specifications), and at which stage in the item processing model is the negative interaction mitigated?

Response processes can be evaluated qualitatively by interviewing respondents during or after item processing through cognitive interviews or think-aloud protocols (Padilla & Lieghton, 2017). Self-reported surveys or observations have been conducted to examine differential anxiety, interest, and motivation based on item features (Leighton, Tang, & Guo, 2017). Additional sources of evidence are uniquely available in digital environments such as biometric data, log-file data, response time, eye tracking, and knowledge tracing, and have been used in various studies to examine test takers' responses processes. However, to meaningfully investigate accessibility, response process data needs to be interpreted in light of the test taker's personal characteristics and the item and/or test design features under which the data were collected.

Understanding Accessibility of the NAEP using Response Process Data

When using the NAEP response process data to understand accessibility, researchers need to clearly specify the claim for which they are gathering evidence. The theoretical and/or evidentiary rationale is needed that links the response process data with the item processing model and implications for accessibility. Research propositions need to account for test takers' personal characteristics, variability in item and/or test design features, and the conditions under which data are available (between and within test takers).

Key questions to consider include:

1. What is the counterfactual? For Claim 1, is the comparison between test takers meaningful and reasonable? Is there an appropriate comparison group? For Claim 2, is there an appropriate intra-individual counterfactual?
2. Which variables will serve as proxies for test takers' personal characteristics? What is the strength of the theoretical rationale linking the variable with the personal characteristic?
3. On which item or test design features is differential inter- or intra-individual processing hypothesized? Is there a meaningful and reasonable data source that can serve as a proxy for item processing?

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.

Ketterlin-Geller, L.R. (2016). Understanding and improving accessibility for special populations. In A. Rupp & J. P. Leighton (Eds.), *Handbook of Cognition and Assessment* (pp. 198-225). Hoboken, NJ: Wiley-Blackwell.

Ketterlin-Geller, L.R., Crawford, L., & Huscroft-D'Angelo, J. N. (2014). Screening to assign accommodations: Using data to make decisions. *Learning Disabilities: A Multidisciplinary Journal, 20*(2), 61-74.

Leighton, J. P., Tang, W., & Guo, Q. (2017). Response processes and validity evidence: controlling for emotions in think aloud interviews. In A. M. Hubley & B. D. Zumbo (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 137-158). Cham, Switzerland: Springer.

Padilla, J-L., & Leighton, J. P. (2017). Cognitive interviewing and think aloud methods. In A. M. Hubley & B. D. Zumbo (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 211-228). Cham, Switzerland: Springer.

Data-Intensive Approaches for Understanding Test-Taking Behaviors and Performances of Students with Disabilities

Andrew Krumm, Assistant Professor
University of Michigan Ann Arbor

There are multiple benefits that can come from analyzing process data to better understand test-taking behaviors. In order to achieve those benefits, there are important factors to consider when working with previously collected data. Below, I briefly describe each of these factors.

Working with previously collected data

Data generated by digital learning and assessment environments introduce different constraints on how those data can be analyzed and used than from data typically used for measurement and assessment purposes (e.g., Wilson & Scalise, 2016). Typically, objects of measurement like latent constructs are developed using available theory and refined over iterative cycles of data collection and analysis. Evidence-centered design (ECD; e.g., Mislevy, Steinberg, & Almond, 2003) and construct mapping (Wilson, 2005) are formal processes that add structure to theory-data-analysis cycles. Approaches like ECD have been applied to making sense of process data from digital learning and assessment environments (e.g., Mislevy, Behrens, DiCerbo, & Levy, 2012); however, a key constraint on these efforts has been that researchers and developers work with data that are ***already being collected*** by a digital environment. The challenge in working with these data, therefore, involves identifying how ***predetermined data elements*** can be used to measure what students know or can do. Said differently, data from digital learning and assessment environments, themselves, place a considerable constraint on what can be measured.

Figure 1 presents a high-level illustration for how features of ECD can be organized to form a validity argument (Mislevy, 2007). Moreover, this figure demonstrates the logic behind how a primary driver, or latent construct, is thought to contribute to what gets observed, such as correct or incorrect on an assessment item. What is observed can be used as evidence to support a claim with appropriate warrants that often stem from theory or prior research. For example, a student's math ability is the primary driver of getting an item correct, and this observation can support the claim that a student knows a related piece of content with proper theoretical support for the appropriateness of the item and relevance of the content. This idea can easily be expanded to measuring test-taking behaviors. A critical difference, however, is that unlike a situation where an assessment designer develops an item to elicit the desired construct, in the case of attempting to measure test-taking behaviors using previously collected data, researchers work with data *as they are* and need to reverse engineer potential drivers or constructs.

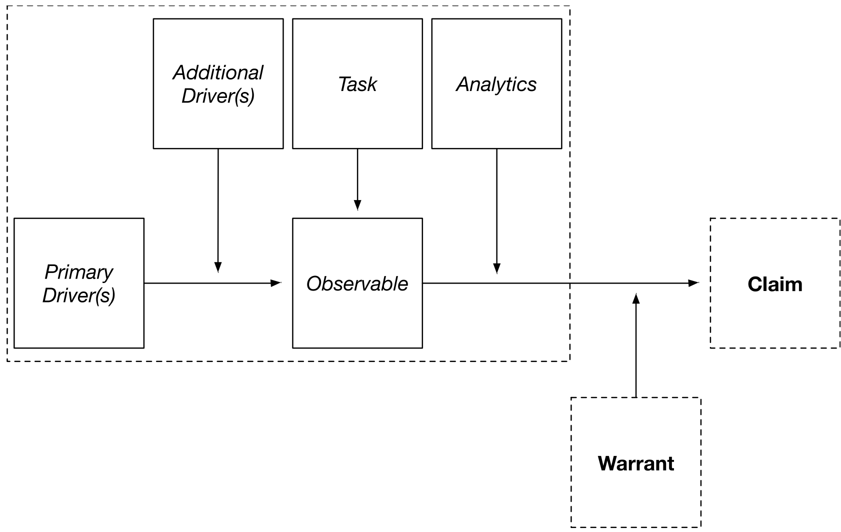


Figure 1. ECD Framework as Validity Argument

Construct validity, predictive modeling, and ECD

Understanding test-taking behaviors using predictive modeling workflows often requires that a name be applied to some form of quantification whether a "predicted value" or "cluster." Critically, naming a set of observations implicates measurement and measurement theory. For example, Baker, Corbett, & Koedinger's (2004) research on the behavior, "gaming the system," which is a form of student engagement within a digital environment where students take advantage of the mechanics of the environment, opened the educational data science field to the possibility of using *predictive modeling* to develop validity arguments related to constructs that were traditionally the purview of psychometricians. A key conceptual difference between predictive modeling and traditional psychometrics is that instead of assuming that observable items are reflective of a latent trait or construct, machine learning-oriented researchers typically begin with the known outcome and work backwards to available data; thus, a primary location for debating construct validity is the manner in which the known outcome was developed or identified. A recent data mining competition demonstrates the potential utility of predictive modeling in understanding if "students spent their time efficiently" on the NAEP using a combination of powerful algorithms and principled variable development (http: tiny.cc/CompAIED).

ECD prompts assessment designers and data analysts to consider other factors that may contribute to what gets observed beyond a primary driver, alone. While what gets observed is critical, how an observation is used to support a claim along with the warrant connecting an observation to the claim (i.e., theory and methods) are all key to understanding the ways in which data on test-taking and test performances are used (e.g., What theories of learning, such as preparation for future learning [Schwartz & Arena, 2013], can help in making sense of students' test-taking behaviors?). Hansen, Mislevy, and Steinberg (2008) outline "ways of thinking, structures for representing knowledge, and tools for supporting reasoning about assessment accommodations" for the NAEP assessment using ECD (p. 100). In particular they highlight the importance of understanding how "additional drivers" influence correct or incorrect observations. These additional drivers serve as alternative explanations for an observation.

Another source of alternative explanations are the "tasks" that are intended to elicit the construct. Both additional drivers and tasks represent primary locations for integrating NAEP process data into claims related to students' performances on test items. Thus, a critical research opportunity is identifying how students avail themselves of various tasks features (specified in both the "observables" and "accommodations" area of the working document) into the modeling process around understanding test performances. *If and when researchers make claims related to how students' use of task features is indicative of some latent construct, then the use of process data shifts from supporting claims related to item performance to supporting claims that a focal latent construct is being measured reliably and well.* This shift, therefore, increases the stakes of the claim one seeks to make with process data from quantifying task features to arguing that process data can be used to measure valued latent constructs.

References

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of Intelligent Tutoring Systems. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, 531–540. Macei´o, Alagoas, Brazil.

Hansen, E. G., Mislevy, R. J., & Steinberg, L. S. (2008). Evidence-centered assessment design for reasoning about accommodations for Individuals with disabilities in NAEP reading and mathematics. ETS Research Report (RR-08-38). http://dx.doi.org/10.1002/j.2333-8504.2008.tb02124.x

Mislevy, R. J. (2007). Validity by design. *Educational Researcher, 36*(8), 463-469.

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and discover in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*(1), 11–48.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement Interdisciplinary Research and Perspective, 1*(1), 3-62.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice-based assessments for the digital age*. Cambridge MA: MIT Press.

Wilson M., Scalise K. (2016) *Learning analytics: Negotiating the intersection of measurement technology and information technology* (pp. 2-23). In: Spector M., Lockee B., Childress M. (eds) Learning, Design, and Technology. Springer.

Mathematical and Non-mathematical Knowledge and Abilities

Steven Ritter, Founder and Chief Scientist-Carnegie Learning

In order to understand what difficulties students might be having in answering NAEP questions, it is important to explore the strategies and approaches that students use to solve problems. This can lead to hypotheses about whether particular strategies favor or disfavor students with disabilities and whether accommodations intended to support these students in fact support the strategies that they use.

Example 1: Pie chart (question 2017-8M3 #2)

Although this was an easy question (see Figure 1) for students (who got 94% correct), I think it provides a good illustration of how strategic problem solving is implicated in answering questions, and it illustrates how difficult it can be to try and isolate mathematical knowledge from other knowledge that may or may not be intended to affect student performance.

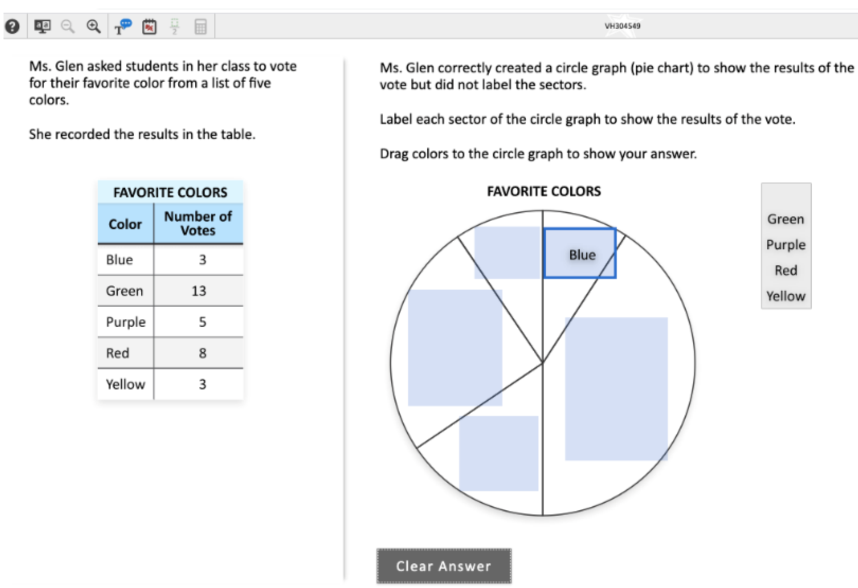


Figure 1: Problem 2017-8M3 #2

The student's task is to label the unlabeled segments of the pie chart, based on the information in the table. The student labels the segments by dragging the labels from the "source" to the right of the pie chart to the chart itself (drop targets highlight when the label is over the chart). Presumably, the mathematical intent of this problem is to see whether the student can demonstrate the equivalence between the table and pie chart representations of this data.

What's interesting to me is the way a student goes about completing this task, and the relation of that strategy to the mathematical knowledge that this question is intending to measure. When I solved this problem, I immediately decided to label the "green" segment first, since it is the largest. A colleague decided to label "blue" and "green" first, because they are the smallest.

The order in which you drag the colors is what I would call a "strategy, and the decision about what strategy to use is partly about mathematics (e.g. being able to tell which is the largest number) and partly about something else (e.g. "do the largest first" may be a consequence of a "do the easiest thing first" heuristic). I can imagine that some students would believe that the appropriate way to solve this problem is to start at the top and go down (so do "blue" first, then "green"). The use of the "go in order" strategy may be a consequence of instruction, even non-mathematical instruction. For example, special education students are often coached to be very systematic in their approaches to problem solving (and test-taking) and so may have a stronger propensity to follow a "go in order" strategy.

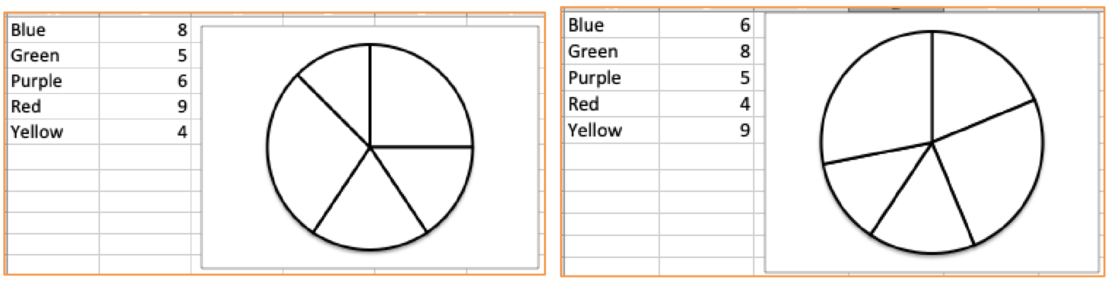


Figure 2: Alternate versions of the pie chart question

Consider Figure 2, which shows (my) alternate versions of this question. It is not hard to imagine a "do the quarter first" strategy for the version of the left, but that is much harder to see as a strategy for the one on the right (which has the same segments in a different order). The point is that subtle characteristics of the question can suggest different approaches, and those suggestions may play out differently for students with disabilities.

These kinds of interactions between student characteristics and the strategies that students choose are worthy of exploration. I believe that the data exists (in "drop choice") to understand these kinds of student strategies.

Example 2: X and Y Intercepts (question 2017-8M3 #4)
This question is a good example of how differences in student mathematical knowledge can interact with question characteristics and phrasing. X and Y intercepts also happens to be an area we recently explored in our MATHia data and with student testing. Among the things we found were that students were much better at determining the y-intercept than the x-intercept (even when the equation was not in slope-intercept form). Students were poor at defining the term "intercept" and often confused intercept with coefficient. The students we tested were proficient at plugging in a zero (or other value) for one variable and calculating the other, implying that, if students' understanding of the term "intercept" were equivalent to "the value of one variable when the other is zero," they'd be able to answer this kind of question using the symbolic form of the equation. Alternatively, students could plug in the coordinates of the given points in the equation to decide which 2 are on the line. This strategy doesn't require any knowledge of what "intercept" means, but it does require an understanding of the relationship

between an equation and a point.

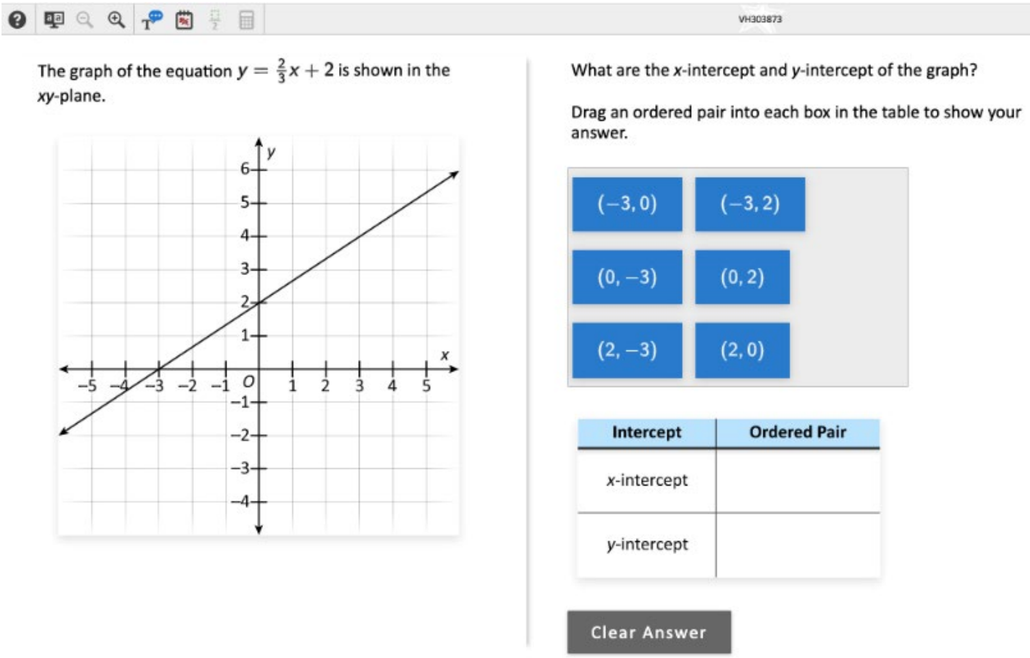


Figure 3: X and Y intercepts question (2017-8M3 #4).

For most students, the equation in this question is a distractor. It is much easier to ignore the equation and look at the graph. A student who understands how to name points on a graph but does not know the meaning of “intercept” (or, perhaps, has a vague notion that “intercept” is sort-of like “intersect”), can identify two significant points on the graph, narrowing the question down to a 50-50 shot at guessing which is the x-intercept and which is the y.

Although the question offers strategies based on the symbolic form of the equation or based on the graph, the graph strategies will be easier for most students. Now consider students with disabilities. It is possible that a student with vision impairment might use increased font size, making the equation more prominent and, perhaps, more likely to be used. In this case, there is little in the data stream that could tell you whether the student used the equation or the graph to answer this question. You might infer that answers which confuse the x and y-intercepts are more common when using the graph.

Teaching Eighth Grade Mathematics to Students with Disabilities

Rajiv Satsangi, Assistant Professor-George Mason University

It is my pleasure to meet and speak with each of you regarding the 2017 Eighth Grade Mathematics National Assessment of Education Progress (NAEP). The following summary highlights central issues associated with teaching eighth grade mathematics curricula to students with disabilities. Issues presented here will hopefully aid you in better understanding testing behaviors, use of accommodations, and outcomes for this population as we collectively work towards improving the academic performance and assessment of students in K-12 education.

Approximately 5-8% of students are identified with or at-risk for learning disabilities, and of this group, nearly 26% struggle in the area of mathematics (Geary et al., 2012). To better understand the characteristics of mathematics disorders, Geary (2003) offered a taxonomy of mathematics disabilities consisting of three subtypes: procedural, semantic memory, and visuospatial. The procedural subtype is characterized by recurrent use of developmentally immature procedures, repeated procedural errors, a lack of understanding of the theoretical ideas related to procedural use, and difficulty sequencing multiple steps in a problem. Age related features represent a developmental delay akin to that of younger children without disabilities; this delay appears to recover and improve across age and grade levels. The semantic memory subtype is characterized by struggles retrieving basic arithmetic facts, a high rate of retrieval errors, and unsystematic reaction times for correct retrieval. Developmental characteristics present themselves through cognitive and performance features that differ from that of younger children without disabilities, and do not change substantively across age and grade levels (McNamara, 2007). This subtype also occurs with phonetic types of reading disabilities. Lastly, the visuospatial subtype is characterized by struggles with spatial representation of numbers and repeated misinterpretation of visually represented material.

A learning disability in these areas can serve as a significant barrier to a students' overall success learning eighth grade mathematics curricula. Students at greatest risk for secondary mathematics struggles are those lacking in counting strategies, numeral identification, magnitude comparison, and working memory (Gersten et al., 2005). Each of these skills are emphasized heavily in the curricular domains of expressions/equations, functions, geometry, and data analysis. Algebra in particular presents significant challenges for students with or at-risk for mathematics learning disabilities. To be successful in algebra, students must first master key prerequisite concepts taught in lower grade levels. For instance, to master symbolic manipulation, students must possess sound arithmetic and computational skills. This is because algebra is considered generalized arithmetic (Carpenter et al., 2003). Moreover, through the transformational activities in algebra, students develop an understanding of equality and equivalence (National Resource Council, 2001). Students strengthen their understanding of relationships between operations and their inverses when engaging in activities focused on reducing expressions and solving complex equations. These tasks require students to consider operations as more than just a set of procedures, but instead as expressions of numerical relationships (Mathews et al., 2012). For instance, when first introduced to linear equations, students must understand the equal sign in an equation to represent a relationship between two numerical values that are equivalent. Developing long-term retention of skills associated with transformational fluency of the equal sign can serve as an obstacle for some students. Despite the initial success that students often exhibit when learning transformational activities, they fail to subsequently recall these same skills a short period of time thereafter (Huntley et al., 2000). Whereas the struggles faced by many students with such memory recall and retention can be traced back to the instructional strategies they engaged in, for students with disabilities, struggles with retrieving procedural information may instead be attributed to their disability (Dunlosky et al., 2013).

Obstacles such as these faced by students with disabilities play a large role in the success they experience in the classroom and may present themselves when students partake in national assessments such as NAEP. However, educators can turn to empirical research to potentially find solutions towards supporting these students on assessments. For instance, research shows students with or at-risk for mathematics learning disabilities learn best when provided explicit instruction, multiple representations of concepts, cognitive strategies (e.g., schematic diagrams to organize information in word problems) and metacognitive strategies (e.g., self-monitoring checklists of steps to complete; Woodward et al., 2012). From this line of work researchers can look to student performance on assessment data to determine trends in student behavior interacting with various types of problems. For instance, what level of engagement do students with disabilities exhibit solving problems accompanied by multiple illustrations and representations depicting values? How were those representations presented with respect to variation in color, size, and the shape of images with accompanying text? Moreover, how explicit were directions and language embedded within the text for students to follow? Findings from such analysis may provide educators with a clearer picture of how students with disabilities approach and engage with test questions and may indicate potential solutions for altering the manner in which we assess mathematics proficiency of diverse learners. Thank you for your time and I look forward to meeting everyone in person later this month.

References

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in the elementary school*. Portsmouth, NH: Heinemann.

Ding, M., & Li, X. (2010). A comparative analysis of the distributive property in U.S. and Chinese elementary mathematics textbooks. *Cognition and Instruction, 28,* 146-180. doi:10.1080/07370001003638553

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-

58. doi:10.1177/1529100612453266
Geary, D. C. (2003). Learning disabilities in arithmetic: Problem-solving differences and cognitive deficits. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 199-212). New York: Guilford Press.
Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, *104*, 206–223.
Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interaction for students with mathematical difficulties. *Journal of Learning Disabilities, 38*, 293–304.
Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of standards-based mathematics education: A study of the Core-Plus Mathematics Project algebra and functions strand. *Journal for Research in Mathematics Education, 31*, 328–361.
Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education, 43*(3), 316–350.
McNamara, B. E. (2007). *Learning disabilities: Bridging the gap between research and classroom practice*. Upper Saddle River, NJ: Pearson Education Inc.
National Research Council (2001). *Adding it up: Helping children learn mathematics*. Kilpatrick, J., Swafford, J., and Findell, B. (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide* (NCEE 2012-4055). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/.

Neurocognitive mechanisms of mathematics learning and implications for investigating NAEP Mathematics process data

Melina Uncapher, Assistant Professor and Director of Education at the UC San Francisco Neuroscape Center and Director of NewSchools' EF+Math Program
University of California San Francisco & EF+Math Program

Davenport's summary in the section on **Student Cognition and Learning in Mathematics** nicely articulates relationships between cognition and behavior that could be investigated with the 2017 NAEP mathematics process data. Here I do not re-state her highly relevant points, but instead focus on a complementary set of testable hypotheses, centered on the ***neurocognitive mechanisms*** that underlie mathematical skill development, which could be further revealed through investigation of relevant 2017 NAEP mathematics 8th grade process data.

'Mathematics networks' in the brain overlap networks that support executive functions and language development. At least two decades of neuroimaging studies investigating mathematics learning have revealed overlapping fronto-parietal networks involved in mathematics, executive functions, and language processing. Does the co-localization of these networks reveal insights into how executive functions and language contribute to mathematics competency? If so, could NAEP data contribute to formulating and testing targeted hypotheses that could lead to high impact interventions, particularly for students with disabilities?

To answer these questions, we need to flesh out a rich and mechanistic understanding of how these higher-order processes relate to each other and how they develop over the course of childhood and adolescence, as these mechanisms will surely point to highly promising areas of investigation. The ultimate outcomes of these investigations should reveal promising and high leverage interventions that can be tested, developed, and implemented for their effectiveness in building mathematical competency in K-12 students. For instance, two areas of parietal cortex--intraparietal sulcus (IPS) and angular gyrus (AnG)--are reliably implicated in mathematical processing, executive functions/learning and memory, and language processing. We provide a high-level summary below:

> Bilateral IPS is consistently implicated in neuroimaging studies investigating how participants represent the abstract meaning of numerosity (e.g., Rosenberg-Lee et al., 2015; Dastjerdi, Ozker, Foster, Rangarajan, & Parvizi, 2013; Kaufmann, Wood, Rubinsten, & Henik, 2011; Nieder & Dehaene, 2009; Ansari, 2008), whereas right IPS seems to code for a topographic representation of numerosity (Harvey, Klein, Petridou, & Dumoulin, 2013). Studies investigating lateralization of IPS activity shows right IPS to be more involved in processing non-symbolic quantities, whereas left IPS is preferentially involved in processing symbolic quantities (Holloway, Price, & Ansari, 2010; Ansari, 2008). IPS is also heavily implicated in top-down, 'goal-directed' attentional control (e.g., Corbetta, Patel, Shulman, 2008) and long-term memory (for review, see Uncapher, Gordon, Wagner, 2014).
>
> Left AnG is involved in the processing of symbolic numbers and arithmetic facts (Grabner,

> Reishofer, Koschutnig, & Ebner, 2011; Price & Ansari, 2011; Holloway et al., 2010; Ansari, 2008), but is also strongly implicated in a language network that supports phonological processing. Interestingly, different subdomains of AnG may support mathematical and language processing: in one study, posterior AnG has been implicated in arithmetical procedures, whereas anterior AnG was more involved in language processing (Andin, Fransson, Rönnberg, & Rudner, 2015). Both areas of AnG have been reliably implicated in executive functions and long-term memory retrieval (e.g., Uncapher et al., 2014).

Areas of promise and opportunity. The foregoing findings raise a field of questions regarding how foundational cognitive processes support mathematical skill development, and the concomitant questions related to how to build those skills in developing brains, particularly for students with disabilities whose disability status is due to executive functioning or language processing disorders. For instance, recent research has revealed that the ways in which the IPS is connected (or indeed hyperconnected) to frontal lobe areas may differ in students with mathematical learning disabilities or developmental dyscalculia (MLD/DD) than in typically developing students (Jolles, Ashkenazi, et al., 2016; Rosenberg-Lee et al., 2015).

A precise, mechanistic understanding of how mathematical skills develop in children and adolescents will require a multi-modal approach that will include behavioral data, cognitive data, neuroimaging data, survey data, and many other forms of qualitative and quantitative data. While NAEP data cannot yet directly inform neural mechanisms of mathematics learning, rich process data from NAEP assessments in high-N samples will go far toward revealing relationships between mathematics, executive functions, and language processing. Specific examples might be:

> Neuroimaging revealed that the left AnG was involved during the retrieval of known arithmetic facts while arithmetical procedures elicited a distributed frontoparietal network (Grabner et al., 2009), so to the extent that studies of NAEP process data can disentangle retrieval of math facts from procedural fluency, we may be able to design and develop interventions that independently lean on each process as appropriate to student needs.
>
> Other findings suggest that left AnG activity is related to the difficulty of the math problems rather than numerical and arithmetical tasks per se (e.g., Wu et al., 2009), so to the extent that NAEP data can disaggregate math complexity/difficulty from numeracy/other operations, could reveal high promise areas of R&D to support students in general, but particularly students with disabilities.

An understanding of the complex and developing relationships between mathematical processing, executive functions, and language will strongly inform the mechanistic hypotheses that can drive the R&D of evidence-based interventions that could make a huge impact on mathematics learning, particularly for students with disabilities or otherwise marginalized students.

Accommodations for Students with Disabilities

Marcelo Worsley, Assistant Professor of Learning Sciences and Computer Science
Northwestern University

Administration of the NAEP Mathematics test includes many practices and accommodations that align with the recommendations of Universal Design for Learning. Some of the built-in accommodations include the ability to zoom, adjust the contrast, activate text-to-speech and access to on-screen highlighting. These accommodations are complemented by additional services that can be arranged as needed. Some of these additional services include having a scribe, taking the exam in Braille, and utilizing custom equipment. As I reflect on these current capabilities, my thinking gravitates to three main opportunities: 1) studying and growing participant awareness of these accommodations, 2) improving and expanding the types of accommodations that are made available to learners and 3) reconceptualizing how we think about disability.

Studying and Growing Participant Awareness

One of the first questions that comes to mind when considering the relatively low scores that students with disabilities receive on the NAEP, is the origin of this problem. Simply put, the low performance could be a reflection of a non-accommodating testing interface, challenges within how we educate students with disabilities, or both. If the problem is the former, it could be equally as plausible that people are not aware of the accommodations or find them to be ineffective. This point about accommodation awareness makes me think about communities. Geographically speaking, we might find that students who are from the same geographic locations, or school district, for example, might be more likely to have a similar level of awareness of the available accommodations. I would anticipate that such a measure would be easily determined from the collected data. The other dimension of community that comes to mind is the heterogeneity of the disability community. It may be that certain disabilities are receiving ample accommodations while other are not. Furthermore, given the correlations between certain disabilities, it is not uncommon to come across students who need multiple accommodations. Hence, an initial area of inquiry is in studying the extent to which people are aware of these accommodations, partitioned into different slices that represent specific subsets/combinations of communities. Addressing this question could help guide how resources should be directed for growing awareness.

Improving and Expanding the Types of Accommodations

In the previous paragraph, I suggested that one possible problem is that test-takers simply are not aware of the resources available to them. If we rule this out, even for a subset of the population, we must next consider the effectiveness of the existing accommodations. I recently completed several NAEP 8th grade problems. Prior to completing the practice problems, I followed a tutorial on how to use the interface. The tutorial aptly pointed out many of the accessibility features that are built into the digital assessment interface. However, the tutorial was intended for someone with full vision, because at many points during the tutorial, buttons were referred to as "this button," without providing additional supporting information. Equally as important is establishing better page navigation for screen reader or the text-to-speech functionality. Several times I attempted to use navigation hotkeys, but these did not seem to get me to where I wanted to go. A related observation was that some of the accessibility features could only be used one at a time. It was not possible to have the text-to-speech feature working, and also drag and drop the desired numbers or highlight the desired text. This presents a challenge to participants may benefit from multiple accommodations, or who simply want to minimize the extent to which their workflow must be modified.

I position these changes as being centrally about improving or refining the existing accommodation practices. A more complex improvement involves rethinking the affordances of different representations for users. For example, several of the questions that I encountered in the sample materials included graphs, or figures. For the sighted user, the visual information provides cues that are instrumental to solving and interpreting the question. The current platform does not provide an option for these graphs and figures to be represented through other modalities that would be beneficial for someone with a visual impairment. Common approaches include providing descriptive text, which would be a strong first step for the NAEP, but I caution that this is insufficient. Instead, it may be preferred to have tactile, or audio representations of these different graphs to enable a more equitable experience. One approach could be to use tactile feedback in the tablet to let test takers feel the different graphs and figures that they encounter in the test.

This challenge also exists for some of the current accommodation tools. For example, the highlighter is provided so that students can actively take note of what they read or observed within a given passage. However, for students that may take the braille version of the test, what sort of tools are provided so that they can easily highlight the passages, and in such a way that mirrors what sighted people get from color and color contrast? Similarly, we might envision using touch screen technology that can more easily accommodate participants that might have non-traditional touch screen gestures. This is particularly true for people with fine motor impairments, where an attempt to touch a single screen location with their finger may results in their entire palm, or the back of their hand, touching the screen. In many cases, these are technologies that currently exist in research labs, but that necessitate additional development before being made available more broadly. With these innovations in place, it becomes easier to provide resources that move beyond accommodations and get closer to enabling equitable participation.

Reconceptualizing How We Think About Disability

This final section advocates for the adoption of the ability-based design framework (Wobbrock, Kane, Gajos, Harada, & Froehlich, 2011)). Universal Design for Learning has been quite impactful for ensuring that people with disabilities are being accommodated in systems that were not designed for them. The ability-based design framework pushes researchers to think more critically about the underlying assumptions that go into the design of a given system. It also shifts the discussion away from dis-ability, and instead, focuses on ability. To think about what it means to focus on ability, consider what would happen if instead of trying to overcome a shortcoming in someone's ability to see a given representation, we design a solution that builds on their assets or abilities. In the case of the graph mentioned above, the goal need not be having the participant visualize the printed graph, the goal is, instead for them to understand the content of the graph and use it in a meaningful manner.

Concretely, the goal is to refrain from defining people or communities based on what they do not have, and to instead focus on what skills or abilities are present. Additionally, ability-based design intends for platforms to adapt to the needs of the user, as opposed to having the user adapt to the technology. Developing interfaces that adapt to the user's needs are in alignment with a previous concern around student awareness of different accommodations. If the system adapts, there is no longer a need for explicit awareness of all possible accommodations. At the same time, self-adaptive systems can reduce the amount of stigma individuals experience from having to setup or activate a given accommodation. More broadly, reconceptualizing how we think about disability changes our orientation, and, consistent with the comments in the previous paragraph, pushes us to think differently about the desired outcome of a given accommodation. It is one thing to try to correct for a documented physical, or cognitive disability. It is something completely different to leverage a participant's heightened perception in an alternate modality such that they are fully supported as they exercise their own way for sense-making with the content at hand.

References

Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-Based Design: Concept, Principles and Examples. *ACM Transactions on*, *3*(3), 1-36. http://doi.org/10.1145/1952383.1952384