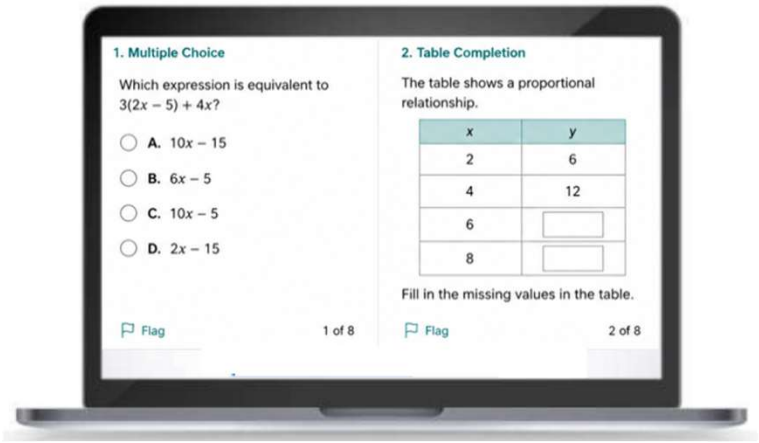
Educational Testing Service

NAEP Usability, Accessibility & Comprehension Research

Presented by Theresa Wilkinson

Education Testing Service

NAEP Usability, Accessibility & Comprehension Research



Agenda

1. Project Overview
2. The Challenge
3. Research Objectives
4. Research Questions
5. Research Approach
6. Key Metrics at a Glance
7. Survey Highlights
8. Key Findings
9. Quantitative Results
10. Recommendations
11. Impact
12. Lessons Learned

Project Overview

What is ETS?

ETS is a global education and talent solutions organization and the world's largest private educational testing organization. ETS develops and administers assessments used by millions of learners worldwide.

Role

- **Lead UX Strategy & Research**

Timeline

- January 2021 – September 2022

Team

- Product Owner
- Business Analyst Leads
- UX Design
- UX Research

Project Overview

The National Assessment of Educational Progress (NAEP) needed to ensure new digital assessment components were usable, accessible, and understandable for students across grade levels and device types before operational deployment.

I led **UX research** supporting multiple large-scale studies focused on **student comprehension, device interaction, assessment usability, and accessibility**. Research findings informed design decisions used by **ETS, NCES, and cross-functional product teams** responsible for future NAEP assessments.

The Challenge

NAEP assessments must work consistently for students of different ages, abilities, devices, screen sizes, and levels of technical experience.

Research was needed to understand how students interacted with **new assessment components**, identify sources of confusion, and ensure designs supported **comprehension** without introducing unnecessary **cognitive burden**.

Research Objectives

- **Evaluate whether students could understand and successfully complete new assessment item types.**
- Identify **usability, comprehension, and accessibility issues** affecting task completion.
- Assess how **device size, screen layout, and interaction patterns** influenced performance.
- Measure **task success, completion time, and user confidence** across different item designs.
- Provide **evidence-based recommendations** to improve question design, instructions, and overall assessment usability.

Research Questions

- Can students **understand the instructions and complete the tasks as intended?**
- What **usability, accessibility, or comprehension barriers** prevent successful task completion?
- How do students interact with the **new item types and interface elements?**
- Does **device size or screen configuration** affect task performance and understanding?
- Which aspects of the design cause **confusion, errors, or increased completion times?**
- What design changes would improve **usability, comprehension, and accessibility?**

Research Approach

- Conducted **moderated usability testing** with elementary and middle school students.
- Evaluated **multiple assessment item types, interaction patterns, and device configurations**.
- Collected **qualitative and quantitative data** including task success, completion time, observations, and participant feedback.
- Identified **usability, accessibility, comprehension, and interaction design issues** affecting performance.
- Collaborated with **researchers, developers, and stakeholders** to review findings and recommendations.
- Delivered **actionable recommendations** to improve assessment usability, accessibility, and student comprehension.

Key Metrics

Student Participants

- 63 students

Number of Studies

- 4 usability studies

Assessment Areas Evaluated:

- Family Structure
- Numeric Entry
- Device-Agnostic
- Adjust Device

Research Methods:

- Usability Testing
- User Interviews
- Accessibility Evaluation

- **25% improvement in comprehension** among 4th-grade students following design revisions.
- **Mixed-methods approach** combining task success, completion time, observations, and participant feedback.
- Identified **usability, accessibility, and comprehension issues** affecting student performance.
- Findings informed **assessment design decisions and future development efforts**.

Key Findings

Finding: Numeric Entry Design Was Generally Successful, but Some Features Added Unnecessary Complexity

Participants: 25 students (4th and 8th grade)

- Students understood **numeric-entry tasks and validation messages**.
- Students understood the **limited math keyboard** and were not confused by disabled buttons.
- Most students preferred their **physical keyboard**, though some found the limited math keyboard helpful.
- Limited math keyboard functionality was well understood and **supported task completion**.
- **Auto-populated commas** created confusion and raised questions about scoring.
- **Negative toggle functionality** was rarely discovered and provided little value.
- Findings informed recommendations to **simplify interactions and improve validation messaging**.

Key Findings

Finding: Family Structure Questions Were Generally Understood, but Terminology and Household Concepts Created Confusion

- Students living in multiple households experienced less confusion with **Variant B than Variant A**.
- Students preferred the **task-oriented progress bar** over the simple progress indicator.
- Students were generally able to identify **family members, caregivers, education levels, and employment information**.
- Confusion was concentrated around terms such as **"worked for pay," caregiver education, and multi-home living arrangements**.
- Students often preferred **clicking rather than dragging** family members between homes.
- Most usability issues involved **terminology, instructions, and interaction patterns** rather than an inability to complete tasks.
- Students reported few concerns about answering family structure questions and generally found the experience **easy to complete**.

Key Findings

Finding: Device-Agnostic Design Requires Balancing Accessibility, Readability, and Consistency Across Devices

- Device compatibility involved more than **screen size**; resolution, input methods, content rendering, and accessibility requirements all influenced the student experience.
- Recommended minimum device specifications included an **11.6-inch screen** and **1366 × 768 resolution** to support usability and accessibility across school devices.
- Smaller screens created **usability and accessibility challenges**, particularly for younger students and those with vision or dexterity limitations.
- Excessive scrolling increased **cognitive load** and could negatively impact reading comprehension and task performance.
- **Horizontal scrolling** was identified as particularly problematic for students.
- **Letterboxing** helped maintain consistency across devices but reduced usable screen space and introduced challenges for some users.
- **Zoom** improved readability but often increased scrolling and interaction complexity.
- Experts emphasized the importance of **readable text, scalable layouts, keyboard accessibility, assistive technology support, and consistent content rendering** across devices.

Key Findings

Finding: Touchscreen Interactions Created Usability Challenges for Students

- Students frequently **mis-tapped small targets or controls**.
- **Drag gestures** were misinterpreted or dropped before reaching the final position.
- Younger students relied on **trial-and-error**.
- Students used **workarounds instead of intended gesture paths**.

Quantitative Results

Numeric Data

Family Structure

- **36 students** participated (Grades 4, 8, and 12).
- **18 students lived in one home; 18 lived in multiple homes.**
- **Progress indicator noticed:** 74% (14 of 19) for the progress bar vs. 23% (3 of 13) for the simple indicator.
- **Click vs. Drag:** 76% (13 of 17) clicked in Version A; 100% (10 of 10) dragged in Version B.

Numeric Entry

- **Time-on-task:** 13 seconds to 1:19.

Quantitative Results

Numeric Data

Device Agnostic

- Conducted research with **8 internal ETS accessibility, usability, and assessment experts**.
- Recommended minimum device specifications of:
  - **11.6-inch screen size**
  - **1366 × 768 resolution**

Support for keyboard, mouse, trackpad, and touch input modalities

- Evaluated device-agnostic considerations across **multiple device types**, including laptops, Chromebooks, iPads, and touch-enabled devices.
- Research included considerations for **Grades 4, 8, and 12 students**.

Adjust Device

- Students frequently **mis-tapped small targets or controls**.
- **Drag gestures** were misinterpreted or dropped before reaching the final position.
- Younger students relied on **trial-and-error**.
- Students used **workarounds instead of intended gesture paths**.

Recommendations

- Use **plain language and age-appropriate terminology** to reduce confusion and improve comprehension.
- Design interactions that **minimize cognitive load** and support successful task completion.
- Prioritize **accessibility** by supporting diverse devices, input methods, screen sizes, and assistive technologies.
- Reduce **unnecessary interaction complexity** and provide clear instructions and feedback.
- Validate designs with **representative students** early and throughout development.
- Continue **usability testing** to identify issues related to comprehension, navigation, accessibility, and interaction design before implementation.

Impact

- **Research-informed revisions improved student comprehension by 25%** in the Family Structure study.
- Provided **evidence-based recommendations** that informed assessment design decisions across multiple studies.
- Identified **comprehension, usability, accessibility, and device-related barriers** affecting student performance.
- Helped stakeholders balance **measurement goals** with the needs of diverse student populations.
- Contributed to improvements in **question wording, instructions, response formats, and device requirements**.
- Generated **actionable findings** used to support future assessment development and validation efforts.
- Findings informed **design recommendations that reduced barriers** for younger students and students with accessibility needs.

Lessons Learned

- **Readability** should be evaluated early and throughout the design process, not after content has been finalized.
- Stakeholders often underestimate the impact of **reading level** on student performance and assessment validity.
- **Small wording changes** can significantly improve comprehension without changing what an assessment measures.
- **Accessibility, readability, and usability** are closely connected and should be evaluated together.
- Testing with **representative students** is essential because expert reviews alone may not identify comprehension barriers.
- Research findings are most effective when they not only identify user challenges but also help stakeholders understand why **readability and comprehension** matter.

Contact

Theresa Wilkinson

Email: theresaw@columbus.rr.com

LinkedIn: linkedin.com/in/theresa-wilkinson-231196

Portfolio: www.w-edge.com/portfolio.html