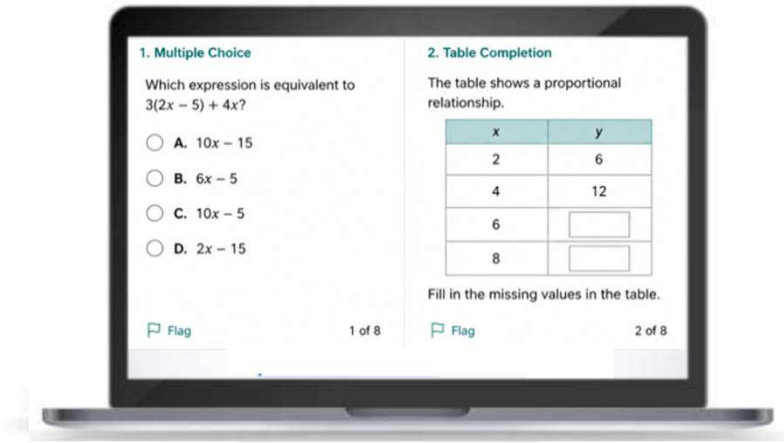
Quantitative Usability & Accessibility Research for NAEP Digital Assessments

Presented by Theresa Wilkinson

ETS Case Study



Agenda

- Overview
- Studies Conducted
- Summary
- Contact

Overview

ETS/NAEP

Educational Testing Service (ETS) develops large-scale educational assessments, including the National Assessment of Educational Progress (NAEP), often referred to as "The Nation's Report Card." NAEP measures student performance across the United States and must provide accessible, consistent assessment experiences for students with diverse abilities, devices, and learning needs.

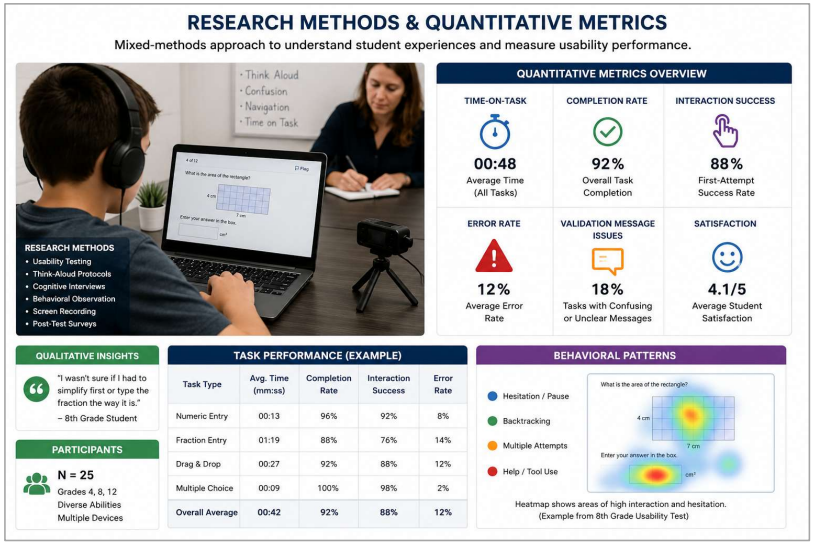
Overview

My Role & Responsibilities

As Lead UX Strategy & Research, I conducted usability, accessibility, and interaction research across multiple NAEP digital assessment initiatives for grades 4, 8, and 12. The work focused on improving comprehension, reducing cognitive load, optimizing touchscreen and keyboard interactions, and supporting device-agnostic assessment delivery across laptops, Chromebooks, and tablets.

Research included usability testing, cognitive interviews, think-aloud protocols, behavioral observation, and quantitative analysis to identify usability barriers and improve assessment experiences for students across a wide range of accessibility and device conditions.

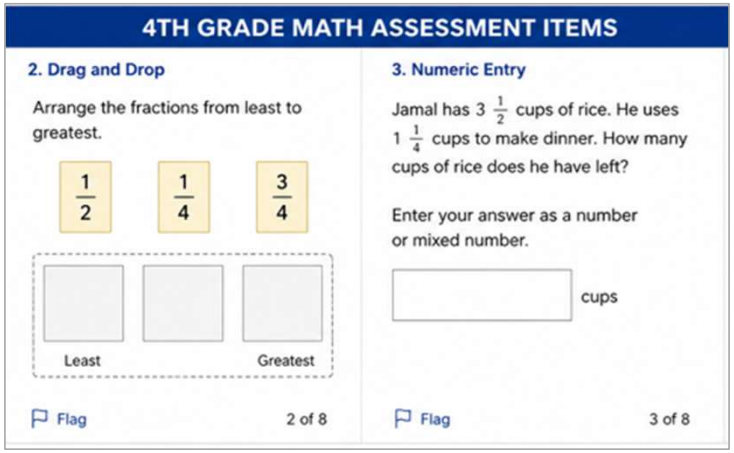
Overview



Research Methods & Quantitative Metrics

Research included usability testing, cognitive interviews, think-aloud protocols, behavioral observation, persona development, and quantitative usability metrics such as time-on-task, completion rates, interaction success, validation-message analysis, progress tracking, and behavioral-pattern evaluation.

Studies Conducted



Studies Conducted:

- **Numeric Entry & Math Keyboard Usability** — Evaluated numeric entry, fraction input, validation messaging, and math keyboard interactions for grades 4 and 8 students.
- **Device-Agnostic Assessment Research** — Examined how NAEP assessments function across laptops, Chromebooks, iPads, and varying screen sizes, resolutions, and input methods.
- **Family Structure Usability Testing** — Tested comprehension, navigation, drag-and-drop interactions, progress tracking, and questionnaire usability for students across multiple household structures.
- **Adjust Device Touchscreen Usability & Interaction Research** — Evaluated touchscreen gestures, drag interactions, touch precision, scrolling behavior, and multi-step task interactions across grades 4 and 8.

Studies Conducted

Numeric Entry & Math Keyboard Usability

Evaluated how students in grades 4 and 8 interacted with numeric entry fields, fractions, validation messaging, and math keyboard interactions. Usability testing with 25 students showed that most students preferred the physical keyboard for numeric entry, while the math keyboard was more useful for fractions and mixed numbers.

Quant results included:

- Average task times from 13 seconds to 1:19
- Physical keyboard dominance for numeric-entry tasks
- Validation issues around commas and formatting
- 8–13% of students reporting usability-related difficulty.

Findings informed improvements to numeric entry, validation messaging, fraction input, and accessibility guidance.

Studies Conducted

Device-Agnostic Assessment Research

Examined how NAEP assessments function across laptops, Chromebooks, iPads, varying screen sizes, resolutions, input methods, and accessibility conditions. Research focused on scrolling, zoom, readability, touch targets, keyboard accessibility, letterboxing, and cognitive load.

Findings identified risks with smaller screens, horizontal scrolling, touch precision, and low-vision accessibility.

Results informed an 11.6" screen and 1366 × 768 resolution baseline, scalable layouts, keyboard accessibility standards, reduced scrolling recommendations, and future Reading and Math framework decisions.

Studies Conducted

Family Structure Usability Testing

Tested how students in grades 4, 8, and 12 completed household composition, caregiver identification, education, and employment questions across multiple household scenarios.

The study compared two questionnaire variants and evaluated comprehension, navigation, drag-and-drop interaction, progress tracking, and completion time.

Quant results showed:

- 100% drag-and-drop success in Variant B
- 74% progress-bar recognition vs. 23% for the simple indicator
- 10% faster completion times, and a 25% improvement in 4th-grade comprehension.

Findings informed improvements to questionnaire structure, terminology, progress tracking, interaction design, and accessibility guidance.

Studies Conducted

Adjust Device Touchscreen Usability & Interaction Research

Evaluated how students in grades 4, 8, and 12 interacted with touchscreen gestures, drag interactions, scrolling behavior, touch precision, and multi-step tasks.

Research included cognitive interviews, think-aloud usability testing, behavioral observation, and screen-recorded interaction logging.

Findings showed:

- Frequent mis-taps, drag precision issues, gesture confusion, repeated interaction attempts
- Cross-grade differences in touchscreen behavior.

Recommendations included larger touch targets, simplified gestures, clearer feedback, responsive layouts, and improved touchscreen interaction standards.

Summary

This ETS/NAEP research program combined usability testing, accessibility research, behavioral observation, persona development, and quantitative usability metrics to improve digital assessment experiences for students across grades 4, 8, and 12.

Key outcomes included improved comprehension, faster task completion, stronger interaction success, clearer progress tracking, improved accessibility guidance, and recommendations that informed future NAEP Reading and Math assessment framework decisions.

Contact

Theresa Wilkinson

Email: theresaw@columbus.rr.com

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

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