Pepe Delgado, Matthias Blaue
EDUC 590
Supervisor: Zarifa Zakaria

UX Case Study

1) Executive Summary

Maslera AI has great pedagogical potential to bridge the gap between declining literacy rates (National Center for Education Statistics [NCES], 2025) and high teacher workloads (National Center for Education Statistics [NCES], 2017, 2022; Steiner et al., 2025) through automation of the creation of tiered reading passages and lesson plans. However, this case study reveals that UX friction and trust deficits burden the platform's value proposition. While the platform functions technically, structural dis-connects force users out of their natural workflow which may lead to task abandonment.

Through moderated usability testing with four secondary teachers, triangulated with a WCAG audit and quantitative analytics, we identified 64 important behavioral data points. The key findings include a severe misrepresentation between the Library and Assignment interfaces forcing redundant navigation loops, a failure of AI simplification to meet pedagogical expectations, and a reliance on static tutorials that fail to provide contextual guidance. Furthermore, analytics reveal a critical 66% non-retake rate among users within a 30-day window, which indicates that initial task completion masks long-term adoption failure.

To resolve these friction points, we recommend five strategic interventions, prioritizing the following three:

1. **Unified Workspace:** Merge the Library and Assignment interfaces to eliminate redundant navigation loops and task abandonment.
2. **Persistent Metadata:** Ensure selected contexts (e.g., text, reading level) persist across the workflow to eliminate redundant data entry.
3. **Fidelity Controls & Co-Pilot Model:** Introduce a "Fidelity Slider" to give teachers control over text simplification. This shifts the AI from a static generation tool to a co-creation workspace. (Additional interventions include replacing static tutorials with in-context onboarding and standardizing minor terminological inconsistencies.)

Strategic Impact

The implementation of these fixes is estimated to save 3 minutes of structural friction per assignment build. Annually, this returns 10.8 hours, or roughly 20% of usable prep time, to every educator for pedagogical planning and differentiation. By removing these blockers, Maslera AI can guarantee the active utilization that is required to protect initial district investments and secure long-term license renewals.

Pepe Delgado, Matthias Blaue
EDUC 590
Supervisor: Zarifa Zakaria

2) Product & Context

According to the National Center for Education Statistics (NCES, 2025), the average reading score on the National Assessment of Educational Progress (NAEP) was lower in 2024 than in 2019, while teacher workloads remained high at approximately 53 hours per week (NCES, 2017, 2022; Steiner et al., 2025). Maslera AI is an educational platform that is built for K-12 teachers and students and designed to address this problem with the automation of the creation of tiered reading passages, materials, and standards-aligned lesson plans.

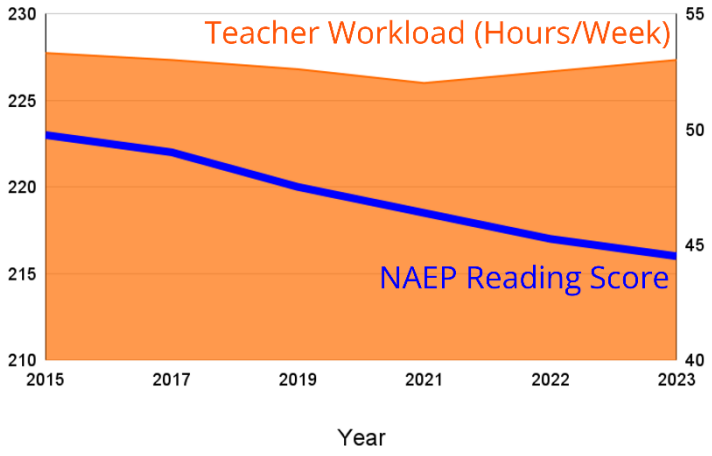

Figure 1: *Declining NAEP Reading Scores vs. Stagnant Teacher Workload (2015–2023).*

The core user flow requires teachers to select ELA standards, generate a customized reading passage, and assign it to their class. The system includes an AI chatbot named "Lex" that provides reading assistance to students and collects and interprets student data for teachers. However, the reality of teacher workflows indicates that this automation is only viable if the interface is frictionless and the AI is completely transparent regarding academic and pedagogical integrity.

3) Stakeholder Ecosystem Map

Maslera AI operates in an educational ecosystem where direct users are influenced by external regulatory mandates and local administrative environments. Mapping these relationships is important, because a failure to accommodate the constraints of any single tier can complicate the adoption of Maslera. The ecosystem has three distinct stakeholder levels:

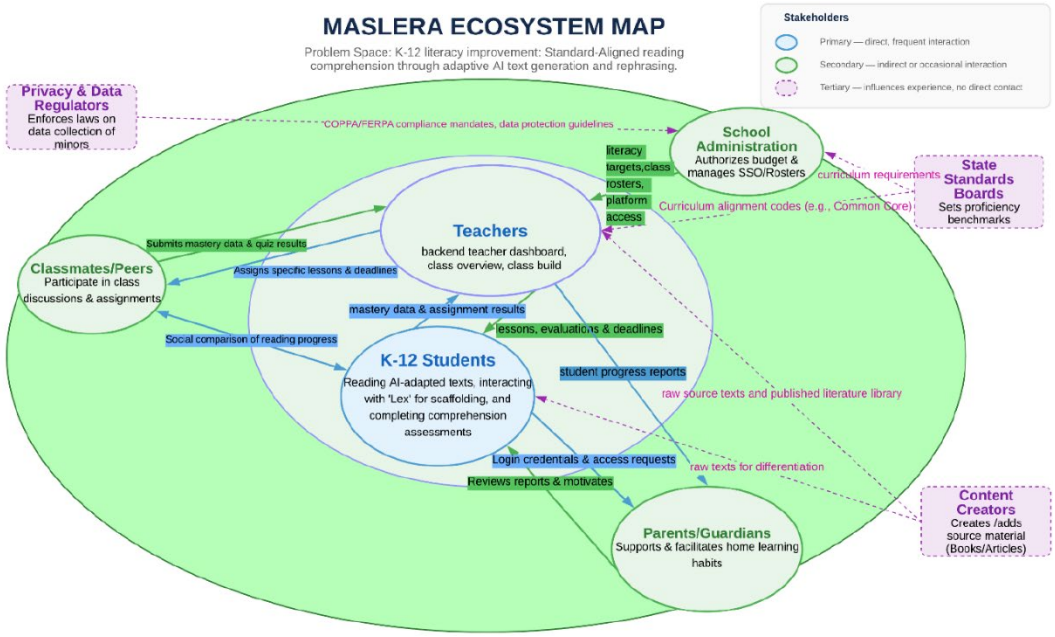

Figure 2*: Maslera AI Ecosystem Map.*

- **Primary Stakeholders**: Teachers (backend class building and assignment) and K-12 Students (reading adapted texts and completing assessments).
- **Secondary Stakeholders**: School Administration (budget authorization and literacy targets), Parents (home learning management), and Classmates (peer environment).
- **Tertiary Stakeholders:** State Standards Boards (proficiency benchmarks), Content Creators (source material), and Privacy/Data Regulators (compliance).

While Maslera's interface is directly used by the primary stakeholders. Its overall adoption relies on satisfying the demands of the secondary and tertiary tiers. When the platform's interface fails to accommodate administrative mandates and grounded classroom realities simultaneously, friction occurs. Our analysis identified three primary UX tensions emerging from these needs:

UX Tensions:

1. **State Boards vs. Teachers:** State boards mandate curriculum compliance codes, but teachers require descriptive standards to translate these into assignments. If the system displays only codes without descriptions, teachers have to leave the platform to search state databases, which breaks the user flow and slows down lesson planning.
2. **School Administration vs. Teachers:** Administrators approve budgets and require data to prove students meet targets. Teachers on the other hand need fast and specific data to support individual instruction. A dashboard that only shows generic data tables fails both groups. Teachers cannot adjust their instruction, and administrators cannot justify the monthly license.
3. **Content Creators vs. Lex AI:** The AI relies on its LLM programming, original texts, and theories implemented by content creators to generate lesson plans. The conflict centers on academic attribution. If the chatbot omits citations for its lesson frameworks, teachers lose professional trust in the software because they cannot verify the content's validity.

4) Learner Journey Map

To situate our analysis in the actual user experience, we revised the initial learner journey map to reflect our usability data. This journey follows a teacher who utilizes Maslera AI to construct a standards-aligned reading assignment and lesson plan. The user path assumes a seamless progression from login to assignment deployment, but the observed journey is characterized by workflow interruptions and cognitive overload. The following stages map the locations where user experience conflicts with the teacher's intent:

- Stage 1 - Discovery & Entry: The user navigates the landing page, identifies pricing, and successfully utilizes Single Sign-On (SSO) for fast entry.
- Stage 2 - Onboarding: The user encounters a mandatory 13-slide tutorial that darkens the background UI. The user bypasses this out-of-context documentation, leading to downstream navigation failures and a reliance on trial-and-error.

- Stage 3 - Assignment Creation (The Representational Divide): The user attempts to select a text and assign it but is blocked by the structural division between the "Library" and "Assignments" tabs. This forces an average of 7-11 redundant navigation steps. Upon finally initiating an assignment, system context does not persist, forcing the user to re-select the book and re-enter grade-level data.
- Stage 4 - Curriculum Planning: The AI successfully lowers Lexile levels but fails to meet secondary academic rigor. The user notes that the simplified text strips out inferencing demands and loses the author's original voice, preventing them from utilizing it without manual revision.
- Stage 5 - Review & Reflection: The user interacts with the Lex Coach to verify pedagogical theory. Pre-update, the AI relies on "general consensus" and refuses citations, resulting in a Severity 3 trust deficit. (Note: A live platform update resolved this by providing explicit literature recommendations, instantly restoring trust).

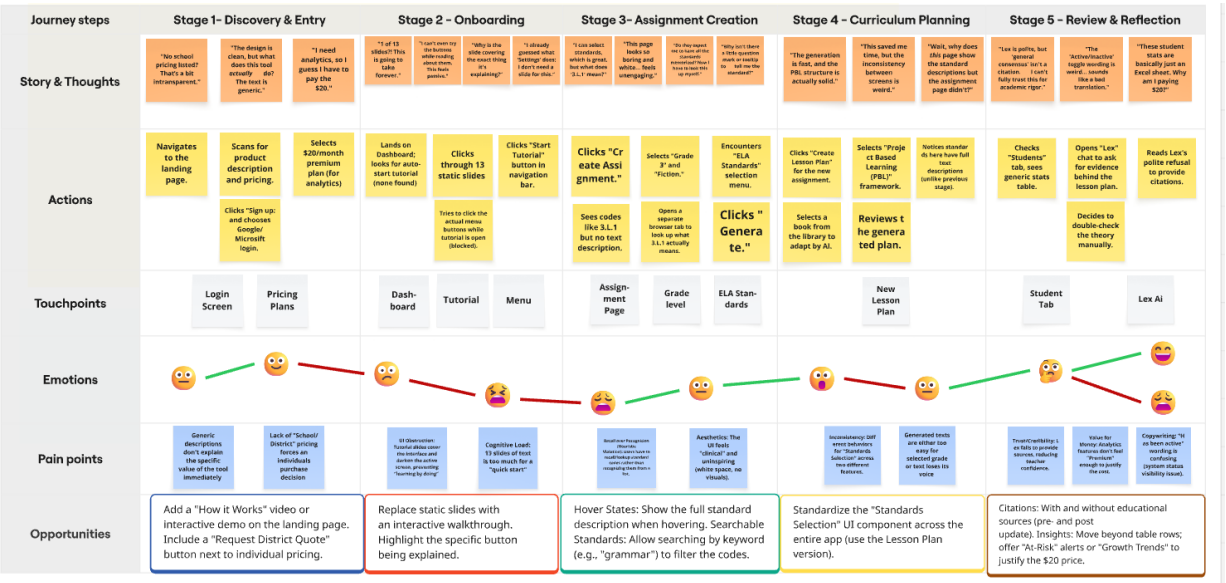

Figure 3: *Updated Journey Map.*

This revised journey illustrates that users might drop off not because of a lack of pedagogical value, but because of frictions in the user experience. Every redundant click, forced data re-entry, and opaque AI interaction reduces the teacher's limited planning time. To detail and explain these journey interruptions, we examine the foundational heuristic and accessibility audit findings.

5) UX Audit Findings

Before we conducted usability testing, we performed a UX audit with the help of Nielsen's 10 Usability Heuristics and WCAG accessibility standards. The goal of this evaluation was to identify barriers embedded in the platform's architecture. The audit revealed that Maslera's interface frequently prioritizes technical backend structures over the user's natural cognitive workflow. We identified the following critical violations:

1. **Opaque Standards Selection (Heuristic Violation: Recognition over Recall):** The assignment generation features display only curriculum codes (e.g., "3. L.1.a") without the descriptive text. This forces users to memorize codes or open external browser tabs to search state databases. This increases cognitive load and reduces planning time.
2. **Obstructive Onboarding (Heuristic Violations: User Control & Freedom; Help & Documentation):** Upon login, users are confronted with a 13-slide static tutorial that darkens the active UI. The system forces a passive and out-of-context reading sequence that acts as a barrier to entry.
3. **Lack of Evidence-Based Citations (Heuristic Violation: Match Between System & Real World):** The Lex AI chatbot provides vague "general consensus" claims when explaining the pedagogical reasoning behind its generated lesson plans. It refuses to provide specific citations and damages professional trust. It forces teachers to manually verify the instructional strategies.
4. **WCAG Accessibility Failures:** The dashboard passed visual requirements for alt-text (1.1.1) and touch target sizes (2.5.8). Keyboard navigation protocols (2.1.1) failed by trapping users within the search bar, which makes the page unresponsive to standard keyboard shortcuts. Error identification (3.3.1) failed because no visible warnings was provided when required form fields were missed, and the "Assign to Class" button was silently disabled instead.

While these findings indicate structural flaws, an expert audit alone cannot measure the actual cost of these barriers. To understand how these technical issues translate into active frustration, workflow abandonment, and the 66% non-retake rate, we developed a usability study to observe these friction points.

6) Study Plan

The UX audit identified theoretical friction points within Maslera's architecture. To move beyond expert evaluation and to identify why usability metrics behave as they do, we used a triangulated, mixed-methods approach.

Participant Selection & Testing Format

We recruited four teachers with established experience in digital lesson planning. This demographic ensured that participants had the pedagogical content knowledge that is necessary to evaluate the AI's output. We conducted moderated, remote usability testing and used a think-aloud protocol to capture feedback.

Task Design

Instead of testing isolated features, participants executed five core tasks that formed a real-world instructional scenario. The sequence included: Initial Entry, Assignment & Standards Alignment, Text Simplification Review, Lesson Generation, and Academic Verification. This structure allowed us to observe natural user behavior and friction, specifically the Task Success and Happiness (user attitudes and trust) dimensions of the HEART framework.

Data Synthesis

The testing sessions yielded 64 qualitative behavioral data points, which we systematically coded using an inductive affinity mapping process and a 15-label codebook. To ensure these qualitative insights were contextualized within platform usage patterns, we triangulated our findings against 66 quantitative user records provided by Maslera AI and the preceding WCAG audit.

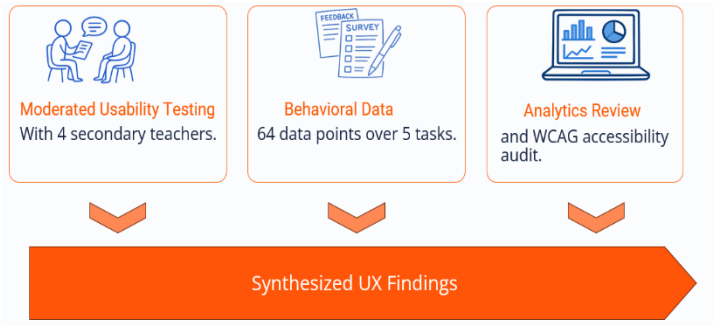

Figure 4: *Triangulated Mixed-Methods Research Design.*

This synthesis of behavioral, analytical, and accessibility data provides the foundation for the detailed usability findings and strategic recommendations that follow in the next sections.

7) Usability Test Findings

The following section presents the primary usability findings from qualitative testing sessions. The data reveals a clear tension between the platform's strong potential to support administrative tasks and the friction created by its current information architecture. Findings are evaluated using Nielsen's 10 Usability Heuristics and prioritized on a severity scale from 1 to 4.

Finding 1: Disconnected Workflow for Text Selection

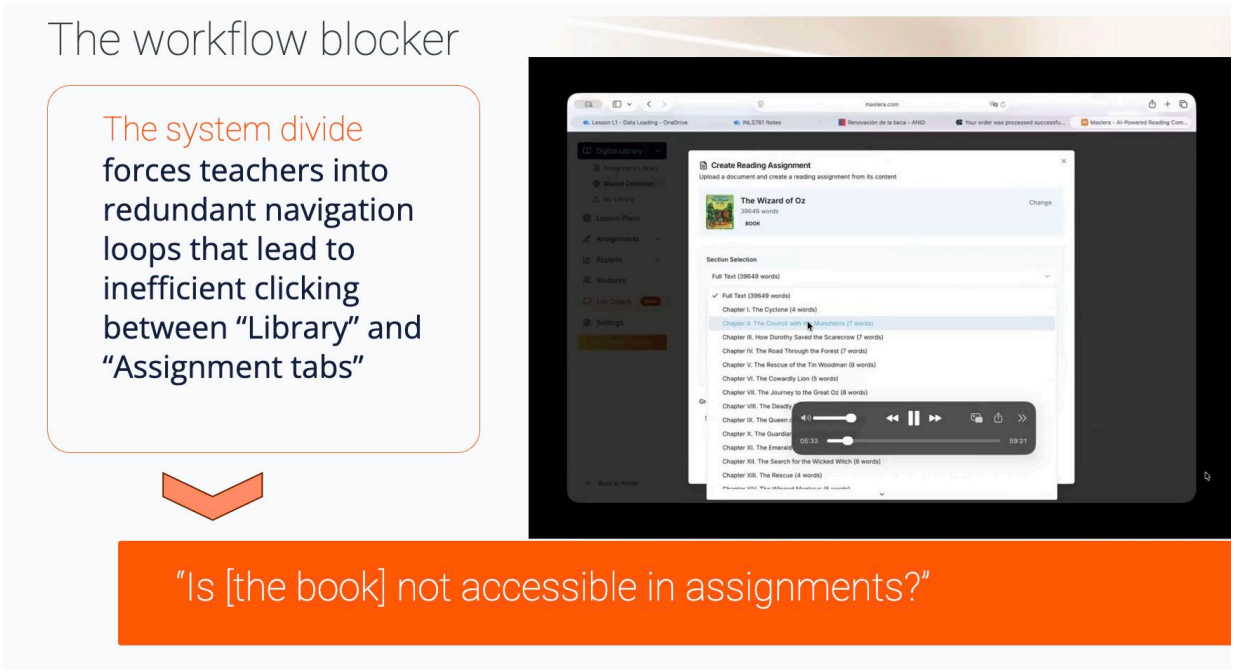

Figure 5: *The Platform Divide*

The data shows a clear mismatch between the system's structure and how users expect to work. Participants tried to move from selecting a text to assigning it in one continuous flow, but the separation between the "Library" and "Assignments" tabs interrupted that process. This suggests that the interface reflects backend organization rather than the teacher's planning workflow.

- Observed Behavior: Participants engaged in verbalized trial-and-error, clicking through the Assignments module, failing to find texts, and expressing verbal confusion. P2 explicitly stated their frustration, noting they had to leave their workflow to find the book collection because it was "not accessible in [Assignments]". All 4 participants leaned closer to the screen and narrowed their eyes, and the narrative paused for almost a minute until they continued the task.
- Relevant Heuristic: Heuristic #2: Match Between System and the Real World (The system workflow does not match the real-world pedagogical workflow).
- Nielsen Severity Rating: 3. Structural disconnect forced users to abandon their intended task path to locate required materials. Not catastrophic, but it significantly impedes workflow and causes frustration.
- Participant Frequency: 4 of the 4 participants experienced this friction.

Finding 2: Redundant Data Entry During Assignment Creation

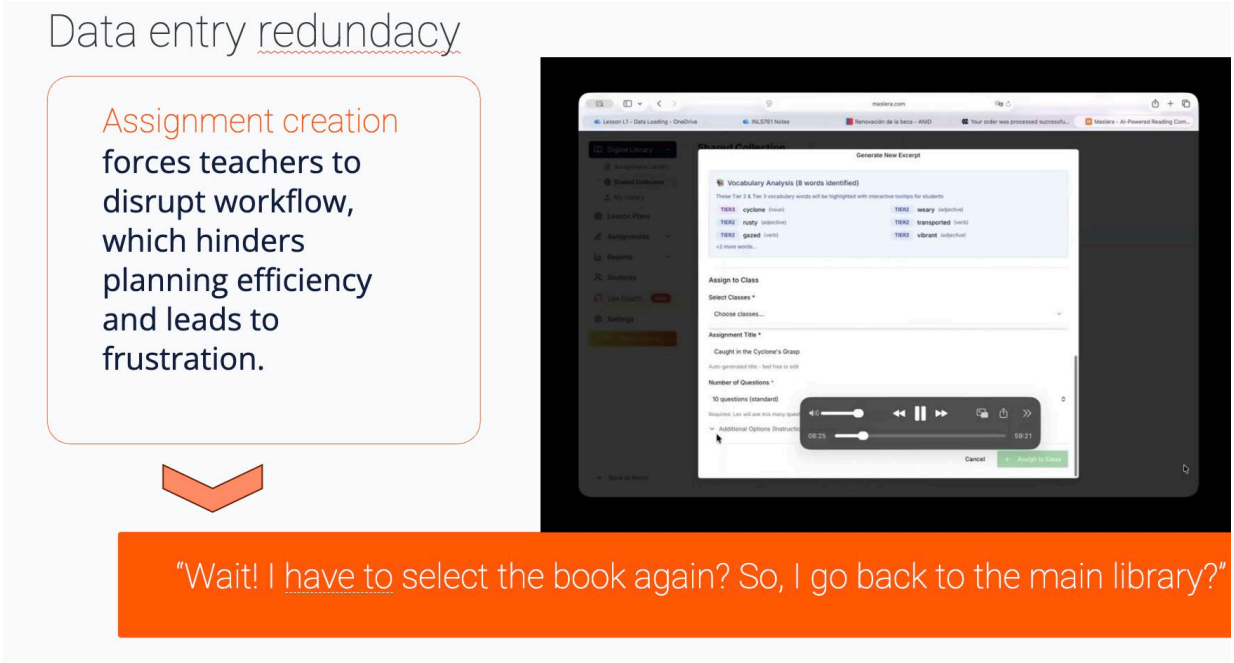

Figure 6: *Redundant Assignment Creation Workflows*

The data shows that the user interface frequently forces redundant actions. Once participants finally located a text and initiated an assignment, the context did not persist.

- Observed Behavior**:** After navigating a specific book page to create a lesson, participants discovered the system requires them to make the same selection again inside a smaller scrollable window. P1 verbalized confusion and her continuous narration breaks, P2 verbalized annoyance when forced to re-enter data, asking, "Wait, I have to select the book again?" and stating that this was "very annoying", while shaking their heads and furrowing their brows.

- Relevant Heuristic: Heuristic #7: Flexibility and Efficiency of Use (The UI requires redundant actions).

- Nielsen Severity Rating: 3. Forcing users to re-enter data, they just selected causes of unnecessary repetition and active verbal annoyance. It degrades efficiency and workflow fluidity.

- Participant Frequency: 4 of the 4 participants experienced this redundancy.

Pepe Delgado, Matthias Blaue
EDUC 590
Supervisor: Zarifa Zakaria

Finding 3: AI Simplification Fails Grade-Level Complexity Standards

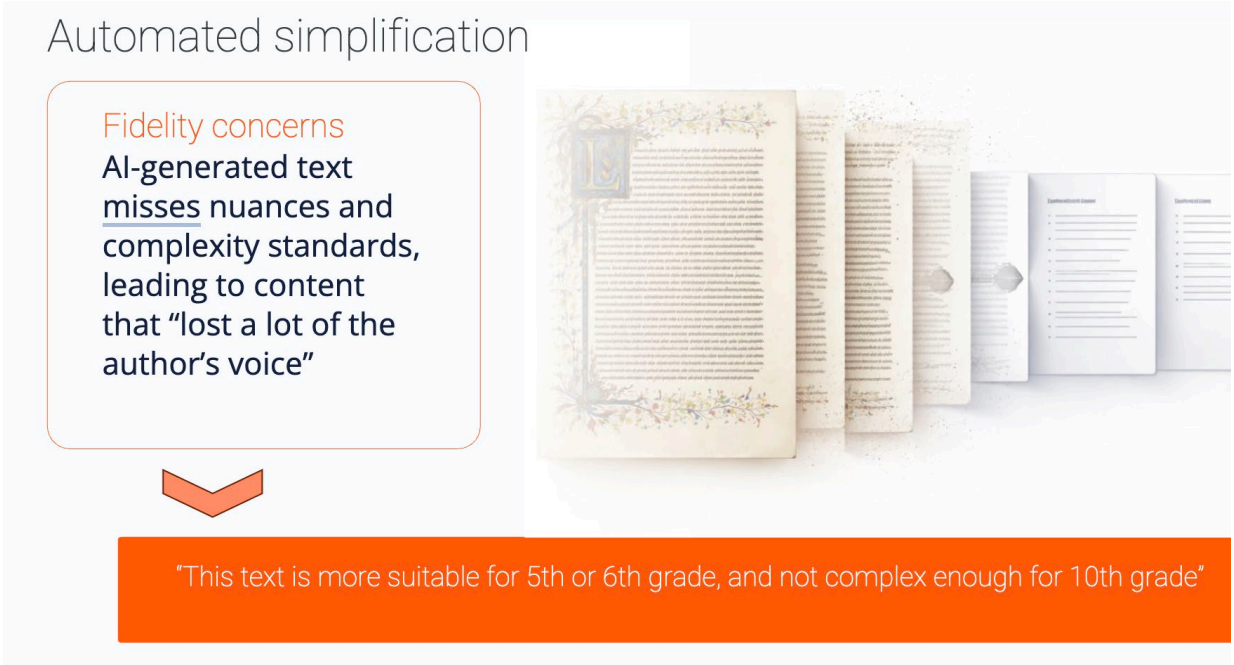

Figure 7: *AI Simplification*

The analysis indicates that automated text simplification heavily prioritizes basic readability over academic rigor. This impeded professional trust for several reasons.

- Observed Behavior: When reviewing texts generated for a 10th-grade level, participants actively critiqued the output. P2 noted it "lost a lot of the author's voice," while P1 evaluated the text as a "more suitable passage for like a 5th or 6th grade classroom," pointing toward a systemic failure to meet the complexity required for secondary education. The participants discussed that topic with a slight smile and noted that they would not use the text as it is for 10th grade.
- Relevant Heuristic: Heuristic #4: Consistency and Standards (The AI fails to adhere to established educational standards).
- Nielsen Severity Rating: 3. Not meeting the intended grade level undermines the tool's primary value proposition and creates a trust deficit. Teachers cannot reliably use the output in a classroom without alteration.
- Participant Frequency: 3 of 4 participants (P1 and P2) explicitly flagged this.

Finding 4: Inconsistent Transparency in AI Citations

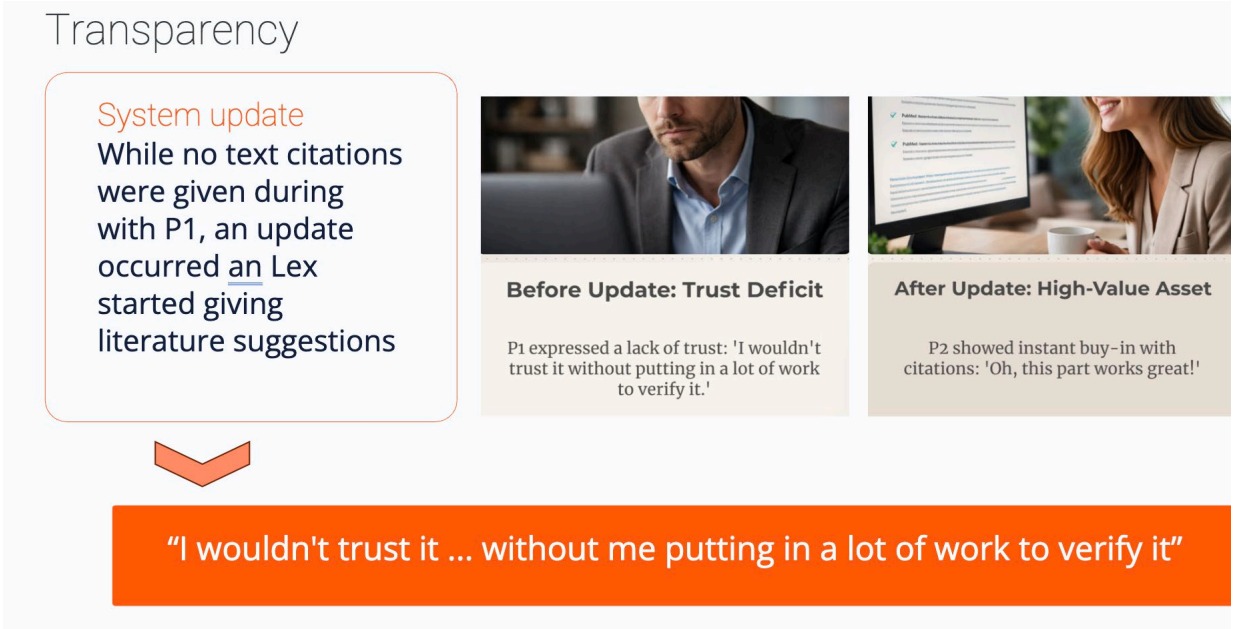

Figure 8: *Platform Update*

An unexpected finding emerged regarding the platform's transparency, likely due to a live system update deployed during the testing window.

- Observed Behavior: In all observations leading up to and during the testing of P1, the AI failed to provide explicit citations, leading P1 to state they "wouldn't trust it ... without me putting in a lot of work to verify it". However, between the testing of P1 and P2, an unannounced platform update likely occurred. For P2, Lex AI effortlessly provided literature recommendations. P2 immediately gave positive feedback and stated with nods and a smile, "Oh, this part works great," instantly transforming the feature into a high-value asset.

- Relevant Heuristic: Heuristic #1: Visibility of System Status (Regarding the provenance and transparency of AI-generated data)

- Nielsen Severity Rating: 3 (pre-update baseline) / 0 (post-update resolution).

- Participant Frequency: 4 of 4 participants experienced this specific variable (P1, 3 & 4 experienced the failure; P2 experienced the successful update).

Finding 5: Onboarding and Knowledge Transfer Failure

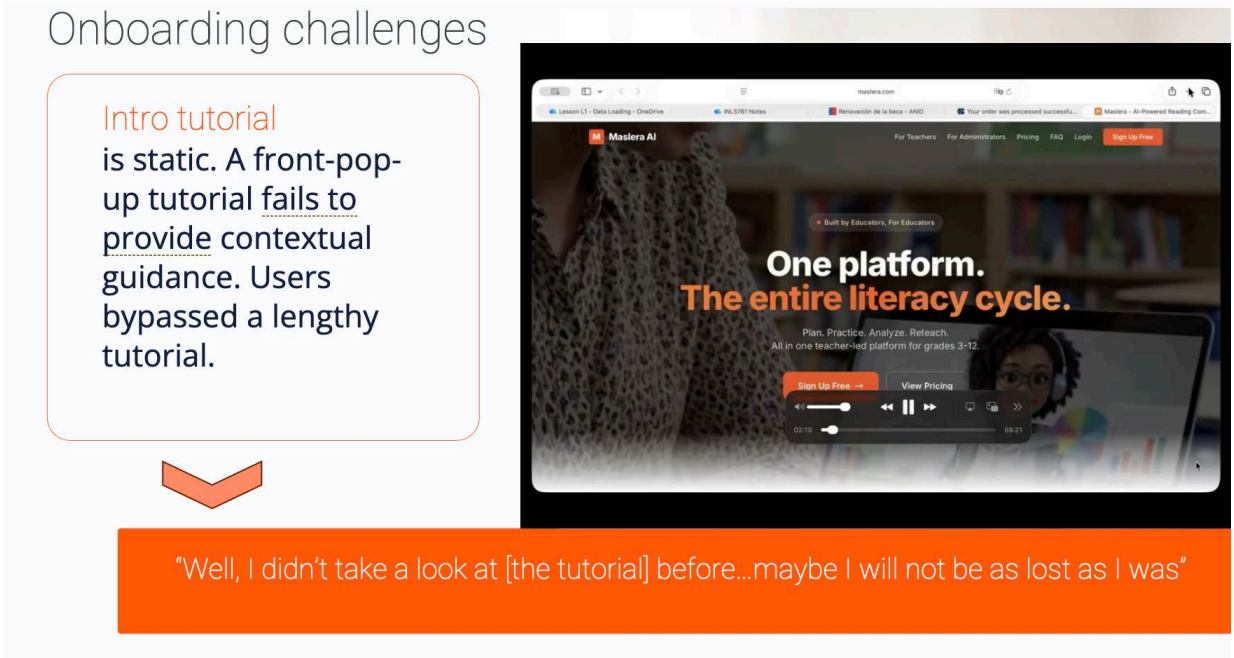

Figure 9: *Onboarding Tutorial*

The data indicates that the platform's onboarding materials (13 static slides) are easily bypassed and completely decoupled from the user's active workflow, leading to downstream task failures.

- Observed Behavior: P3 skipped the tutorial and had severe problems navigating the dashboard. Rather than blaming the system's usability, P3 explicitly verbalized this knowledge gap, stating, "well, I didn't take a look [at it] before ... maybe I will not be as lost as I was". This points to a reliance on out-of-context memorization rather than in-the-moment guidance. Additionally, P1 and P2 read through the tutorial but immediately after the completion, they had to orient themselves inside the dashboard again before being able to continue the task. P3 mentioned that it is a lot of information to get through.

- Relevant Heuristic: Heuristic #10: Help and Documentation (Help should be integrated, contextual, and focused on the user's immediate task, rather than requiring upfront, out-of-context viewing).

- Nielsen Severity Rating: 2.

8) Analytics & HEART Metrics

To evaluate Maslera's effectiveness as an instructional tool, this case study applies Google's HEART framework: Happiness, Engagement, Adoption, Retention, and Task Success. Each dimension is tied directly to observed quantitative data to assess whether the platform supports meaningful learning.

Perceived Value and Trust)

Retakes: 33% yes, 66% no, Completion: 86%

Interpretation: Most users complete assignments, but few return to revise. This suggests that feedback does not motivate improvement. If users found value in the system's guidance, retake rates would be higher. As a result, users complete tasks, but do not trust or value the feedback enough to revisit them.

Engagement (Depth of Interaction)

Time on task: 0 to 28 minutes, average 21.03 minutes; Retakes: 33%

Interpretation: Time on task varies widely. Some users spend significant time, while others disengage quickly. Combined with low retakes, this shows that interaction is not iterative. Therefore, engagement is inconsistent and lacks depth.

Adoption (Initial Use)

Completion rate: 86%,

Interpretation: Users can navigate Maslera and finish tasks. As a result, adoption is strong at the surface level.

Retention (Sustained Use)

Retakes: 33%; Paused assignments: 8 users; Score range: 0% to 90%

Interpretation: Few users return to improve their work. This can be evidenced when paused sessions suggest drop-off before completion or revision. Overall, Maslera does not support sustained or repeated use, and the score range varies widely.

Task Success (Learning Outcomes)

Comprehension levels: 4 Mastered, 20 Near Mastery, 9 Developing, 13 Need Support

Interpretation: Performance is uneven. A large portion of users fall below mastery. Therefore, completion does not lead to consistent understanding. Likewise, task completion does not equal learning success.

HEART Synthesis

Across all metrics, a consistent pattern emerges. Users can start and finish tasks, but they do not return, deepen engagement, or achieve consistent outcomes. Maslera's system supports access, not progression. Overall, the system works as a task completion tool. It does not yet function as a system that supports revision, reflection, or learning growth.

9) Triangulated Insights

This report synthesizes qualitative usability tests, accessibility audits, and quantitative analytics to evaluate the Maslera AI platform. Qualitative findings from four teachers show that automated text simplification provides clear value, but a separation between the Library and Assignments interfaces creates significant workflow friction. The study also identifies a trust deficit tied to unclear AI-generated citations, though a mid-study update shows that trust can improve when transparency is addressed. Quantitative data reveals uneven engagement. Metrics such as score, comprehension level, and completion status show that users interact with the platform, but they do not indicate where or why breakdowns occur. These metrics capture outcomes, not process. They do not explain user decisions, sources of confusion, or moments of friction, and therefore cannot fully represent the user experience.

Pattern 1

Triangulating the audit report with behavioral data reveals a clear mismatch between system structure and user expectations. The audit identifies a “Representational Divide” between the Library and Assignment modules. The usability testing shows the consequence. Teachers are forced into redundant navigation loops, which disrupt their workflow. The audit explains where the system is fragmented. The behavioral data shows how that fragmentation creates cognitive overload. Participants pause, search, and express frustration. These moments help explain the drop-off patterns seen in analytics. The quantitative data reinforces this pattern. Irregular performance and uneven task completion suggest that users do not move through tasks in a smooth or consistent way. However, the metrics alone cannot explain why. When combined with qualitative findings, it becomes clear that workflow disruption drives these inconsistencies.

Pattern 2

A second pattern emerges when comparing onboarding to actual use. The audit confirms a 13-slide tutorial, but usability testing shows that participants skip this material. Later, they attribute task failures to their own “lack of preparation.” This suggests that static, front-loaded instruction does not support real use. Users need guidance within the workflow, not before it. The issue is not access to information, but timing and context. Without embedded support, users must rely on memory or guesswork during tasks. The quantitative data reinforces this gap. Time on task varies widely, yet it does not predict performance or reveal where friction occurs. Some users spend more time without better outcomes, which suggests that effort is spent navigating confusion rather than completing tasks efficiently.

Pattern 3

Finally, triangulating the audit’s focus on AI trust with the qualitative results reveals a clear pattern in adoption. The audit raised concerns about how missing citations might affect professional buy-in. The usability findings confirm this. Before the update, the “black box” nature of the AI created a Severity 3 trust deficit. After citations and academic links were introduced, the same feature shifted into a high-value resource for teachers.

This shift shows that trust directly shapes adoption. When teachers understand and verify AI outputs, they are more willing to rely on them in instruction. The quantitative data adds another layer. It shows wide variation in performance and engagement, from high mastery to students needing support. However, these metrics do not explain the cause of that variation. They do not show why users pause, disengage, or perform poorly. When combined with qualitative findings, it becomes clear that trust and transparency play a key role in these outcomes.

10) Recommendations

Based on usability findings, ecosystem tensions, and evidence synthesis, the recommendations below aim to reposition Maslera from an AI content generator to an instructional co-pilot that supports teacher decision-making, preserves pedagogy, and improves workflow efficiency.

1. Unify the Instructional Workflow (High Priority)

Teachers move between the Digital Library and Assignments with repeated steps and lost context. Given this friction, it is recommended to combine text selection, assignment creation, and configuration into one continuous workspace. By allowing teachers to select, preview, and assign texts in a single flow with key inputs, such as text, grade level, and standards across steps, the platform will remove the need to repeat selections. This aligns the interface with how teachers plan instruction. It reduces cognitive load and speeds up task completion. Success can be measured by a reduction in time-on-task for assignment creation and a decrease in redundant navigation clicks between tabs.

2. Restore Teacher Agency in Questioning (High Priority)

The fact that teachers cannot edit AI-generated questions limits control and reduces trust. By adding tools to edit, delete, regenerate, and create questions, Maslera will provide multiple AI-generated options by type, such as comprehension, analysis, or inference. Also, allowing teachers to tag questions to standards will shift AI from a replacement to a collaborator. Therefore, teachers stay in control of instruction. Success can be measured by tracking the utilization rate of the new editing and regeneration tools, indicating that teachers are actively exercising their instructional agency rather than accepting the default AI output.

3. Introduce Text Fidelity Controls (High Priority)

AI simplification reduces rigor and alters author voice. By adding a fidelity control with clear options such as “Preserve Original,” “Balanced”, and “Simplify”, Maslera could show side-by-side comparisons between original and adapted text. Even further, explaining what changed and why could help teachers balance accessibility with rigor to protect the integrity of literary texts. Success can be measured by a higher percentage of teachers directly assigning the AI-generated text without exporting it for manual revision.

4. Make AI Transparent by Default (High Priority)

Inconsistent citations and unclear outputs reduce trust. By attaching sources, citations, and generation logic to all AI outputs, teachers would understand where content comes from before they use it. Transparency builds confidence and adoption. Thus, including a “how this was generated” view could ensure consistent behavior across sessions. Success can be measured by an increase in positive qualitative feedback during subsequent usability testing regarding teacher trust in the platform's academic rigor.

5. Upgrade Analytics to Be Actionable (Medium Priority)

Current dashboard shows data but not decisions. If Maslera shows skill-level breakdowns such as inference, vocabulary, and evidence of use, the platform could highlight priority of students and suggest the next steps. By adding trend views over time, teachers would not have a set of clear actions they can take in class. Success can be measured by an increase in the assignment retake rate, indicating that teachers are effectively using the data to intervene and re-assign tasks.

11) Limitations & Next Steps

This case study offers insight into usability, trust, and workflow challenges within Maslera, but several limitations shape how the findings should be interpreted. The sample size was small (n=4) and limited to ELA teachers, which restricts generalizability across subjects, grade levels, and broader educational contexts. The study also captures first-use interactions, so many friction points may reflect initial learning rather than persistent usability issues that emerge over time.

Because sessions were moderated and used a think-aloud protocol, participant behavior may have shifted due to observation. Users may have persisted longer than they would in a natural setting. The use of structured, scenario-based tasks also limits validity, as these tasks may not reflect real classroom constraints or real decision-making conditions.

A live update of Maslera during testing, related to AI citation transparency, introduced variation across participant experiences and reduced consistency. While the platform quantitative analytics provided performance snapshots, they lacked explanatory depth and did not reveal the causes behind user behavior or outcomes.

The study also does not measure instructional impact or student learning gains. It does not include a comparison to other tools or traditional practices. These constraints position the findings as exploratory that point to key patterns in usability and trust rather than offering conclusive evidence of overall platform effectiveness.

References

Interaction Design Foundation. (2015). *Google's HEART framework for measuring UX*. https://ixdf.org/literature/article/google-s-heart-framework-for-measuring-ux

National Center for Education Statistics. (2017). National Teacher and Principal Survey (NTPS), 2015–16. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/surveys/ntps/

National Center for Education Statistics. (2022). National Teacher and Principal Survey (NTPS), 2020–21. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/surveys/ntps/

National Center for Education Statistics. (2025). National Assessment of Educational Progress (NAEP): Reading assessment. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/nationsreportcard/reading/

Steiner, E. D., Woo, A., & Doan, S. (2023). All work and no pay—Teachers' perceptions of their pay and hours worked: Findings from the 2023 State of the American Teacher Survey (Research Report). RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1108-9.html

Pepe Delgado, Matthias Blaue
EDUC 590
Supervisor: Zarifa Zakaria

Figures