USC Rossier School of Education
EdPolicy Hub

USC Dornsife Dana and David Dornsife College of Letters, Arts and Sciences
Center for Applied Research in Education

From Proof of Concept to Platform: Testing the RPPL Visualizer for *Eureka Math*^{2} Implementation Health Data

University of Southern California

University of Pennsylvania

Great Minds, LLC

Riverside Unified School District

May 2026

Executive Summary

Many factors beyond the quality of any given set of instructional materials affect curriculum implementation in K–12 mathematics, as classroom instruction sits at the intersection of teacher knowledge and belief, school culture, leadership strengths, district policy, and professional learning, among other influences. Understanding how those forces interact with student outcomes requires data systems that go well beyond a single research study. Yet in practice, most implementation monitoring relies on ad hoc analyses that end when a grant ends.

This white paper documents a six-month, AIMS-funded "proof of concept" project in which our team—the EdPolicy Hub and Center for Applied Research in Education (CARE) at the University of Southern California (USC), Great Minds (GM), Riverside Unified School District (RUSD), and the University of Pennsylvania (UPenn) —explored whether Research Partnership for Professional Learning's web-based "Visualizer tool" could serve as a durable platform for GM's envisioned Eureka Math^2 (EM2) Implementation Health Dashboard. Developed with AIMS funding, the Visualizer is a web-based tool designed to help practitioners, product developers, and researchers make timely, research-aligned use of curriculum implementation and professional learning data. The Visualizer transforms data collected from common tools, like Google Forms or Qualtrics, into a standardized format. It allows practitioner, researcher, and product-developer audiences to explore collected data through interactive dashboards and visualizations without the need for technical expertise or external analysts.

From October 2025 through April 2026, we worked to integrate collected walkthrough observation and curriculum-embedded assessment data from all 11 RUSD EM2 pilot schools into an interactive, shareable dashboard. We solicited structured feedback from RUSD district and school leaders, iterated on the prototype in response, and systematically documented what worked, what didn't, and what would need to happen next for this platform to realize its full potential.

Our headline finding is cautiously optimistic: adapting the Visualizer to an entirely new implementation context—new data structures, new constructs, and organizations outside the RPPL network—is feasible, and the resulting prototype generated genuine enthusiasm from RUSD district leaders. At the same time, the project surfaced real constraints, particularly around data quality, role-based access control, and the alignment of construct language between researchers and educators. We close with concrete recommendations for AIMS teams considering similar work, and for GM and RPPL as they weigh next steps.

Table of Contents

Introduction: Implementation Data Dashboard Proof-of-concept

Teachers' successful use of curriculum and instructional materials to guide their students' learning depends on many factors beyond the quality of the materials themselves. Some of those factors are intrinsic to teachers, such as their knowledge, experience, perceptions, beliefs, and expertise (Brown, 2009; Kolodner et al, 2003; Rogers et al, 2011). Teachers also regularly adapt and/or supplement their curriculum, which can facilitate or impede student learning (Fogelman et al, 2011).

Other factors, "enablers of and barriers to implementation," are external to teachers, presenting at levels ranging from state and district policies, to school leadership and culture, to availability of high-quality, aligned professional learning supports, to students' experiences with the curriculum (e.g., Darling-Hammond et al, 2024, Elmore, 2004, Fullan & Quinn, 2023). Curriculum providers have begun codifying these factors into structured frameworks connecting implementation conditions to student outcomes. Great Minds' Framework for Implementation Success is one such example.

Researchers and curriculum providers have long collected data on these factors within the boundaries of individual studies or grant-funded projects. The challenge is data collection rarely outlives research project funding, and even when they do, resulting analyses often sit in reports that rarely reach the practitioners who need them most. We asked what it would look like to give district and school leaders an interactive window into their own implementation data, updated over time and linked to student outcomes.

Starting at the beginning of the 2025-26 school year, Great Minds compiled Implementation Health Reports for individual district clients, manually assembling walkthrough data, assessment results, and survey responses into PowerPoint slides. The Visualizer project set out to test whether RPPL's web-based data platform could transform that manual process into something more sustainable, interactive, and scalable—and whether the lessons learned in RUSD could inform similar efforts by other AIMS curriculum providers.

This white paper documents the goals, process, successes, challenges, and implications of our proof-of-concept effort. We write it not as an account of a fully realized system, but as an honest reckoning with what a first-generation prototype can and cannot do, and what it would take to go further.

The Collaborative Team and Our Respective Roles

This project required four organizational partners, each contributing distinct expertise.

With USC Principal Investigator Dr. Anna Saavedra leading the team, USC provided project management, IRB and data use agreement coordination, district interview coordination, and led drafting of this white paper. The USC team—also including Dr. Tiffany Wu and Mr. Marshall Garland—contributed the broader research framing and literature context for the project, drawing on our ongoing AIMS/ENCORE-funded study of EM2 implementation in RUSD.

Great Minds contributed implementation and student outcome data, specifically, walkthrough observation rubric data collected by GM's Customer Success team and field research team, and curriculum-embedded assessment data from EM2's online platform. They also provided subject matter expertise on the EM2 implementation framework throughout the project. Dr. Jordan Vossen (GM research team) was our primary data liaison, who provided updates to Maciej as new rounds of walkthroughs were completed. Dr. Helena Connolly served as the GM senior advisor to the project.

Riverside Unified School District (RUSD) participated in co-design feedback sessions and offered the candid feedback and a practitioner perspective that was most essential to shaping the prototype. Ms. Mary De Guzman, Instructional Services Specialist in Mathematics, is the lead RUSD contact, along with Dr. Valerie Backstrom, Grants and Project Development Manager.

University of Pennsylvania (Dr. Maciej Pankiewicz) created the Visualizer tool and the technical expertise to adapt it to EM2 data. Maciej's previous experience building the original RPPL Visualizer for ELA measurement was indispensable. As he noted more than once during team meetings, this project helped us all appreciate just how much complexity lives beneath the surface of even a "simple" data dashboard.

Our project also benefited from the input of advisor Anna Poole, who brought experience from New York City's large-scale implementation of high-quality instructional materials (HQIM) across multiple math and ELA curricula. Her perspective helped us anticipate platform needs extending beyond RUSD and EM2.

The RPPL Visualizer: Tool and Platform

The RPPL Visualizer is a web-based data platform designed to help practitioners, curriculum developers, and researchers make timely, research-aligned use of implementation data. Developed with AIMS funding, it transforms data collected through common survey instruments—such as Google Forms or Qualtrics exports—into a standardized format and renders that data through interactive dashboards. Pre-configured visualizations include radar charts for construct-level patterns, milestone graphs for time-series comparisons, bar and column charts for cross-school or cross-group comparisons, and stacked distributions for scale-item spread.

Critically, the Visualizer also incorporates AI-generated narrative summaries of the data, automatically synthesizing patterns across constructs and flagging notable results. This feature proved among the most discussed, as stakeholders held varying views on the nature of interpretive work the tool should do for its users versus leaving interpretation to practitioners themselves.

Though our prototype version doesn't currently build out this feature, the Visualizer also supports benchmarking and anonymized comparisons across organizations. We designed it with modularity in mind, allowing teams to define new constructs, upload data in compatible formats, and customize dashboard views. Prior to our project, only RPPL member organizations had used the Visualizer, exclusively with RPPL's ELA measurement toolkit.

Our project represents the first use of the Visualizer as infrastructure for an external research partnership: that is organizations outside the RPPL network, working with new data types, new constructs, and new stakeholder audiences. While RPPL's internal pilots test the Visualizer's functionality with RPPL-generated measures, our project tested something different: whether the Visualizer could function as a generalizable platform that other curriculum providers and research teams could adapt to their own implementation monitoring needs. The short answer is yes—but with meaningful caveats that we address below.

Data Sources: What We Used and What We Hope to Use

The USC research team is collecting a wide range of quantitative and qualitative data measures as part of our primary AIMS project. During the 2025-26, school year, 302 teachers across 11 RUSD schools are using EM2 curriculum, and 499 teachers across 19 schools are using the prior version of the curriculum (Eureka Math).

For this mini-grant project, due to the labor demands of each additional data type, we limited the number to two: 1) a measure of teachers' implementation of EM2 curriculum and 2) a student outcome. For both implementation and outcomes, we chose data sources that GM collects in Riverside Unified, as well as across ~500 of their client districts. We decided on this "universal use" selection criteria so that GM could realistically consider potential expansion from a prototype into a fleshed-out tool. Building the tool using data only USC collects and GM doesn't plan to in the future would have been less useful.

With these considerations, we decided to build the prototype dashboard integrating these two primary data sources, which describe in more detail below.

Walkthrough Observation Data

GM's Customer Success Program (CSP) and field research team both conduct classroom walkthroughs using a shared rubric. Walkthroughs assess four overarching dimensions:

1) Curriculum Use: a teacher's skill in facilitating Great Minds lesson components and pedagogical elements.
2) Student-Centered Instruction: emphasis on students' cognitive lift and discourse.
3) Gathering and Using Evidence: how teachers use curriculum-embedded formative and summative assessments to provide feedback.
4) Student Engagement: participation in collaborative and independent lesson activities.

Each dimension encompasses two subconstructs and observers record both numerical ratings and qualitative notes for each.

GM conducted CSP walkthroughs in two rounds, October through mid-December and February through April, with each observation lasting approximately 15 minutes. GM field research walkthroughs occur three times per year in October, February, and mid-March through mid-April, with researchers typically observing a full lesson. For the purposes of this proof of concept, we used walkthroughs conducted by both CSP and field research team members to increase the sample size for our implementation data source. By April 2026, we had a full year of walkthrough data from October 2025 through April 2026, representing 91 observations across all 11 RUSD EM2 pilot schools (ranging from 5 to 11 observations per school), with 64 observations conducted by CSP team members and 27 by field research team members.

Important caveats: GM in partnership with school leadership did not randomly select which teachers, grades, and schools to observe. Rather, GM and school leaders collaborated to identify which teachers participate in each round, with considerations ranging from scheduling to the professional

trust required for observation to feel safe and productive. The non-random sampling limits what we can conclude from the data, which we address in the Data Quality section below. The observations are also not equally distributed across grades for the reasons stated above.

Curriculum-Embedded Assessment Data

EM2's online platform generates student assessment data through two types of instruments: topic quizzes and end-of-module (unit) assessments. For our prototype, we used assessment data from grades three through six across all 11 pilot schools, representing 2,638 students out of the 7,184 enrolled in TK-6 RUSD schools using EM2. This data captures student performance at the school and grade level. However, its utility is limited by uneven teacher uptake, as not all teachers assign the platform-embedded assessments with equal frequency.

We link walkthrough and assessment data at the school level, which is the most defensible level of aggregation given the sample sizes involved. Teacher-level and student-level aggregation would require much larger and more systematically collected walkthrough samples than we currently have.

Data Sources on the Horizon

Our longer-term vision for the Visualizer Implementation Health Dashboard encompasses additional data sources that fall outside the scope of this mini grant. These include teacher and school leader pre- and post-year surveys addressing enabling conditions, GM coach observation data, EM2 portal click data reflecting teacher usage patterns, and student administrative records including interim and state assessment scores. Comparisons between GM districts would also be valuable. Integrating these additional sources would dramatically enrich the dashboard, though each requires its own data quality control process and visualization logic. We flag them here to make clear the distance between our current prototype and the full Implementation Health Index we ultimately envision.

Building the Prototype: Process, Iteration, and What We Learned

From Synthetic to Real Data

Our team began the project in November 2025 with synthetic (i.e., fabricated) walkthrough data, which allowed the UPenn team to build the core dashboard infrastructure before data sharing agreements with UPenn/RPPL were finalized. This practical workaround proved valuable, letting us move quickly in the early months and giving the team something concrete to react to. Jordan subsequently provided real walkthrough and assessment data in the format GM exports from Qualtrics, which UPenn integrated into the live prototype.

One early technical challenge was that GM's data export format required anonymizing teacher names and manually explaining the file structure before the Visualizer could ingest it. Maciej worked with Jordan to identify an automatable data preparation process, a prerequisite for any system that aims to update dashboards efficiently over time. This is a solvable problem though not a trivial one and represents an ongoing investment of technical time. Future partnerships may want to develop uniform data formatting standards across organizations, incorporating built-in automated data conformity checks.

Structuring Feedback from RUSD

We designed our feedback process around structured conversations with RUSD district and school leaders. In late March 2026, we shared the prototype with Mary and Valerie from RUSD's district leadership team, walking them through the dashboard's current features and asking a set of prespecified questions about audience, utility, and data gaps. Their feedback was detailed, thoughtful, and in several cases pointed to platform limitations that would require significant additional development to address.

Key themes from the March 30 and April 2 feedback sessions included:

1) Audience and access: District leaders and school leaders were the appropriate primary audiences for the dashboard rather than individual teachers. Sharing teacher-level data directly with teachers risked unhealthy competition or shame. Grade-level aggregations would be useful for school leaders and coaches, while district-level aggregations served the district leadership team. Implementing this vision requires role-based authentication, which would be a major backend undertaking beyond our current scope.
2) AI summaries: RUSD leaders strongly supported AI-generated summaries and preferred them to start with a district-wide overview before drilling into school- and grade-level specifics. Mary emphasized that the summaries should highlight not just what the data shows, but what is driving high and low performance—the "why" behind the patterns. The UPenn team iterated on the summary format in response to this feedback.
3) "Bright spots": RUSD leaders wanted the platform to highlight exemplary schools and help other schools learn from what those schools were doing differently. Maciej confirmed this was technically feasible through filtering, aggregation, and AI summary features.

4) Construct alignment: The walkthrough rubric's subconstruct labels—terms like "pedagogical elements" and "lesson component facilitation"—were familiar to GM researchers but unfamiliar to RUSD educators. Jordan agreed to provide plain-language definitions of each construct, which Maciej incorporated as tooltip pop-ups accessible via an icon in the dashboard. This was a workable short-term fix for a deeper, longer-term alignment challenge.

5) Enabling conditions: RUSD leaders noted that some of the most actionable insights in GM's implementation data related not to instructional fidelity per se, but to enabling conditions. They specifically mentioned teacher feedback norms, school leader efficacy perceptions, and the alignment (or misalignment) between how teachers and school leaders assess instructional progress. Though these data types are not currently in the prototype, they represent the most valuable future additions.

Challenges

Technical Complexity and Scope Management

This project was more technically complex than most of us initially anticipated. Maciej's/UPenn team's prior experience building the original RPPL Visualizer meant he was well-positioned to manage the adaptation. That experience also gave him a clear-eyed view of how much work lies ahead. Moving from the current prototype to a full-featured platform with role-based access, multiple simultaneous data sources, and the ability to serve dozens of districts at once would require a dedicated software development effort well beyond the scope of a single mini-grant.

The current prototype is best understood as a proof of concept in the strictest sense: it demonstrates that the core idea is viable and that RUSD leaders find genuine value in visualized implementation data. It does not yet demonstrate that the platform can scale to serve GM's broader client base. Those are different claims, and we want to be clear about which one this project supports.

The Data Quality Challenge: Being Transparent About Limitations

If there is one lesson this project surfaced more forcefully than any other, it is that data quality is the essential prerequisite for any implementation monitoring system, and that current data quality in our prototype is limited in ways that matter.

The fundamental challenge is that our walkthrough data is neither representative nor comprehensive. With 91 total observations across 11 schools, we have a range of five to 11 observations per school. The number of observations allows us to construct school-level averages, but the method for selecting classrooms, which is primarily based on convenience, does not allow us to make inferences about what is driving differences across schools, or whether a particular school's average reflects its teaching population or its selection into the walkthrough sample. This is not a flaw in the platform; it is a constraint of the real-world data collection environment. While RUSD school culture is evolving toward more normalized, less fraught classroom observation, systematic random sampling of walkthrough subjects is not yet standard practice.

Similarly, curriculum-embedded assessment uptake is uneven across teachers, meaning that the student outcome data in our prototype reflects only the subset of students whose teachers consistently assigned platform assessments. This creates a potential confound: schools that appear to have strong assessment outcomes may simply have teachers who are more consistently using the platform, not necessarily teachers whose students are learning more.

We recommend that any team (including ours) using or considering a similar platform build data quality diagnostics directly into the dashboard. Users would benefit from knowing response rates, sample sizes, and potential selection biases as they interpret visualized results. Mary specifically requested that the dashboard flag data quality issues prominently, and Maciej incorporated response rate transparency into the current prototype. Future iterations should go further, providing users with guidance on how to interpret results given data quality constraints.

In the interim, we used synthetic data in early prototype testing precisely to allow us to assess platform usability without being distracted by data quality issues—a methodological choice we recommend to other teams undertaking similar work.

What We Produced: Deliverables and Current Status

As of spring 2026, our collaborative team has produced:

1) A functioning prototype dashboard based on RPPL's Visualizer, incorporating three rounds of GM walkthrough observation data from all 11 RUSD EM2 pilot schools and curriculum-embedded assessment data from 2,638 students in grades three through six.
2) Interactive visualizations including radar charts displaying construct-level walkthrough patterns, bar and column charts supporting school-by-school comparisons, and grade-level breakdowns of student assessment performance.
3) AI-generated narrative summaries of walkthrough and assessment data iteratively refined based on RUSD leader feedback to begin with district-wide overviews and drill into specific patterns.
4) Documentation for all walkthrough construct labels, providing plain-language definitions accessible to non-researcher audiences, is available both in the main documentation and through chart-integrated tooltips for in-context guidance.
5) A mock-up demonstrating how walkthrough data at two time points can be compared using the platform's side-by-side visualization capability.
6) Documentation of our data preparation process, including the Qualtrics-to-Visualizer conversion workflow for GM's walkthrough data format.
7) The platform code, which UPenn/RPPL team will release as open source via GitHub (https://github.com/pcla-code/AIMS-Visualizer), enabling other curriculum providers and research teams to adapt and build on the infrastructure.

We also plan to present the prototype at the June 2026 AIMS Collaboratory convening, where we hope to gather feedback from other AIMS teams and districts about the platform's potential broader utility. Due to schedule constraints, Maciej will not attend in person; we plan to demonstrate the platform via a short video combined with live screen sharing.

Perspectives from Key Stakeholders

RUSD District Leaders: "This Is Work We're Currently Doing by Hand"

Perhaps the most validating feedback we received came from Mary, whose response to the initial prototype demonstration was essentially: this is what we need. RUSD is currently reviewing implementation and outcome data manually, receiving formatted reports from GM, and reconstructing patterns that a well-designed dashboard could surface automatically. The prospect of a living, interactive platform that updates as new data becomes available, and that allows district leaders to ask their own questions of the data rather than waiting for a researcher's analysis, was genuinely compelling.

Mary and Valerie articulated a clear vision of how the platform would integrate into district practice:

- District leaders would use aggregate views to make decisions about district-wide professional learning priorities
- School leaders and coaches would use school- and grade-level views to identify areas for targeted support
- AI summaries would help busy administrators quickly grasp key patterns without having to interpret visualizations themselves.

They were also direct about what the platform should not do: surface teacher-level data to teachers directly, at least at this stage, given concerns about unintended competitive dynamics.

Great Minds: Promising but Not Yet Ready for Broad Adoption

GM's response to the project was thoughtful and appropriately measured. Helena, who led GM's engagement with the project, acknowledged the prototype's promise while also naming the organizational questions that GM needs to resolve before committing to broader investment. What does GM need to learn about the use of our current system-level implementation monitoring tools before investing in new infrastructure? Does the value proposition justify the ongoing cost and technical maintenance? Which data sources should such a system include, and how would GM manage data sharing agreements at scale?

These are not objections to the project; they are the right questions for an organization at this stage of the work to be asking. GM indicated it will make decisions about further investment after the June 2026 AIMS convening, having heard from district partners and seen how comparable organizations are approaching similar challenges.

UPenn/RPPL: From Tool to Platform

Maciej and the UPenn/RPPL team approached this project with both expertise and candor. From the beginning, Maciej distinguished between the Visualizer as a tool—optimized for RPPL-generated measures within the RPPL network—and the Visualizer as a platform—generalizable infrastructure that external research partnerships could adapt. This project has tested the latter for the first time, and the experience confirms both the potential and the difficulty of that ambition.

RPPL's commitment to open sourcing the platform code is a meaningful contribution to the field. It means that organizations with technical capacity to adapt the codebase can do so without starting from scratch—and that the insights from this project, including the Qualtrics-to-Visualizer data preparation workflow and the EM2 construct documentation, are available to future teams.

USC: From Theory to Reality

After viewing a RPPL demo of the Visualizer tool at a previous AIMS convening, Anna conceptualized this proof-of-concept project, drafted the proposal, and convened the team. The USC team's main takeaways build upon those of the other three stakeholders: 1) RUSD would find a developed dashboard valuable, aligning with the originally hypothesized reasons, 2) GM needs to determine whether further investment is a viable strategic direction, 3) UPenn demonstrated the reasons for why the platform concept is technically challenging to fully realize, and 4) the open-sourced platform code, and this white paper, are meaningful contributions.

Implications for Other AIMS Teams and Beyond

We write this section with other AIMS curriculum providers specifically in mind, though the lessons apply more broadly to any team considering building implementation monitoring infrastructure.

Start with the Practitioner Question, Not the Data Source

The most important design decision we made was to ground the prototype in a specific practitioner question: How can district and school leaders use implementation data to make better decisions about professional learning and instructional support? Every subsequent design choice flowed from that question. The data sources we prioritized, the aggregation levels we focused on, the AI summary format we refined, incorporated input from Mary and Valerie about what they needed.

Teams that start with the data they have and work backward to use cases tend to produce dashboards that technically function but that practitioners don't use. The more productive sequence is to identify the decision a practitioner needs to make, then identify the data that would inform that decision, then build the platform to surface that data clearly.

Be Transparent About Data Quality from Day One

Every implementation monitoring system will have data quality limitations including incomplete data, non-random sampling, and uneven response rates. The temptation is to address these issues quietly in methodological appendices while presenting clean, polished visualizations to practitioners. We recommend against this approach. Practitioners who are surprised by data quality limitations after they have started relying on a platform are less likely to trust it. Practitioners who understand the limitations from the beginning can use the data wisely, treating it as directional rather than definitive and calibrating their confidence accordingly.

Build data quality indicators directly into the dashboard interface, make sample sizes visible, and flag constructs where data is sparse. Transparency about the types and quality of data necessary to make valid inferences builds rather than erodes trust.

Plan for the Technical Complexity

We cannot overstate how much backend development is required to build a genuinely multi-stakeholder implementation monitoring platform. Role-based authentication, automated data pipelines, the ability to handle multiple simultaneous data sources in different formats, integration of time-series comparisons—each of these capabilities requires significant programmer time. The current prototype does not have all these features and adding them would be a major undertaking.

Our recommendation: scope your proof of concept carefully, be explicit with all partners about what a first iteration will and will not include and plan the next development phase before the first one is complete. The gap between "we have a prototype" and "we have a platform" is large, and it is not primarily a data gap. It is a software engineering gap.

Align Construct Language Early

One of our more tractable challenges was the mismatch between GM's walkthrough rubric language and the language RUSD educators use to describe instruction. Terms like "pedagogical elements" and "lesson component facilitation" made sense to researchers but were unfamiliar to practitioners. We addressed this through tooltip definitions, which worked as a short-term fix. A more durable solution involves co-designing the construct labels with educators from the beginning. Future iterations of this work should involve teachers and school leaders in naming the constructs they will ultimately be asked to interpret.

Involving Students' Voices

The proposal for this project included student voice as a priority, primarily through exit ticket data that we had hoped to integrate into the prototype. Though we were unable to include this data source in the current iteration because it is not yet part of the EM2 program, we want to reaffirm its importance. Students' own impressions of the curriculum—whether they find the work engaging, whether they feel they are learning, whether the pacing feels right—represent a critical and often underweighted input into implementation monitoring. We encourage future iterations of this work to prioritize student data alongside the teacher- and leader-facing measures.

Consider the Longer-Term Vision for Cross-District Comparisons

One of the most intriguing possibilities that surfaced during our stakeholder conversations was the prospect of benchmarking RUSD's implementation data against other GM client districts, helping RUSD understand whether its implementation challenges are idiosyncratic or common to EM2 districts broadly. This kind of benchmarking could be enormously valuable for both districts and GM as a curriculum provider, and RPPL's Visualizer is designed to support anonymized cross-organization comparisons.

Realizing this possibility would require GM to establish data sharing agreements with multiple districts and to create infrastructure for aggregating and anonymizing cross-district data at scale. It would also require careful conversation with districts about the uses to which comparative data might be put. These are not insurmountable challenges, but they require deliberate planning. They represent a qualitatively different kind of investment than building a single-district prototype.

Next Steps: What Comes After the Proof of Concept

We close with a candid assessment of what the next phase of this work requires, and from whom.

For Great Minds: The central question is whether GM wants to invest in becoming an organization that provides ongoing, data-informed implementation monitoring infrastructure to its districts—not just formatted reports, but a living platform that districts can access and use independently over time. Answering that question requires clarity about the business model (who maintains the platform, who bears the cost), the data strategy (which sources to prioritize, how to ensure data quality at scale), and the organizational capacity to support districts in using implementation data for continuous improvement. The June 2026 AIMS convening is an appropriate moment to gather input from districts and peer organizations, and to listen carefully to what districts say they need.

For UPenn/RPPL: Releasing the platform code as open source is a meaningful contribution, and other teams can learn from what we built through the code. Future funding could support the development of additional constructs, improved data pipelines, or the role-based authentication features that would make the platform genuinely multi-stakeholder.

For the AIMS community: If this project demonstrates anything for other curriculum providers, it is that the challenge of implementation monitoring at scale is a systems problem. Building infrastructure that connects curriculum use to student outcomes, in a form that practitioners can use, requires collaboration across organizations with very different expertise. The AIMS Collaboratory model, which brought together a curriculum developer, a school district, a research university, and a technical development partner, is the right structure for this kind of work. We encourage other teams to look for similar cross-sector partnerships rather than attempting to build implementation monitoring capacity within a single organization.

For USC: We plan to continue documenting implementation patterns in RUSD through our AIMS/ENCORE grant, and to keep the door open for further Visualizer development as GM determines its future direction. The white paper and prototype presentation at the June 2026 convening are the outputs of this mini grant; the ongoing AIMS/ENCORE work is the platform on which future iterations could be built.

Conclusion

When we proposed this project, we described it as a "proof of concept" exploration of whether RPPL's Visualizer could serve as the platform for a GM Implementation Health Dashboard. Six months in, the answer is yes:

- The Visualizer can be adapted to EM2 data
- District leaders find real value in the resulting prototype
- The open-source codebase provides a foundation for future development
- The collaborative process we built, connecting curriculum provider, school district, research university, and technical development partner, is a model worth replicating.

The qualifications are equally real:

- Data quality is limited by non-random walkthrough sampling and uneven assessment uptake
- The platform currently lacks the role-based access control that would make it genuinely multi-stakeholder
- Construct language requires better alignment with practitioner vocabulary
- The distance between a first-generation prototype and a scalable platform requires an investment of technical capacity that no single mini grant can fully fund

We have tried to document both sides of that ledger honestly in this paper. Implementation monitoring and building shared infrastructure that connects curriculum use to student outcomes in a form that practitioners can use is a very challenging task. However, it is valuable in ways that extend well beyond a single research study and our work points to its promise.

We look forward to continuing the conversation at the June 2026 AIMS convening, and to learning what other teams are building as we all work toward the same goal: understanding what is happening in classrooms and using that understanding to improve student learning.

Appendix A: Key Terms and Acronyms

AIMS: Accelerating Implementation of Mathematics in Schools. The funding collaborative supporting this project.

CSP: Customer Success Program. Great Minds' tiered district support program, which includes walkthrough observations conducted by Success Leads.

EM2: Eureka Math^2. Great Minds' K–12 mathematics curriculum, the instructional materials at the center of this project.

ENCORE: The AIMS sub-program funding the broader USC/GM/RUSD research partnership of which this mini grant is a part.

GM: Great Minds. The curriculum developer and project partner.

HQIM: High-Quality Instructional Materials. A term used to describe curricula that have met evidence-based quality standards.

RPPL: Research Partnership for Professional Learning. The organization that developed the Visualizer tool and served as technical partner.

RUSD: Riverside Unified School District. The district partner for this project.

Visualizer: RPPL's web-based interactive data platform, the focus of this proof-of-concept project.

Appendix B: Visualizer Platform Prototype

Figure B1: Main view of the Visualizer interface: implementation/ walkthrough data.

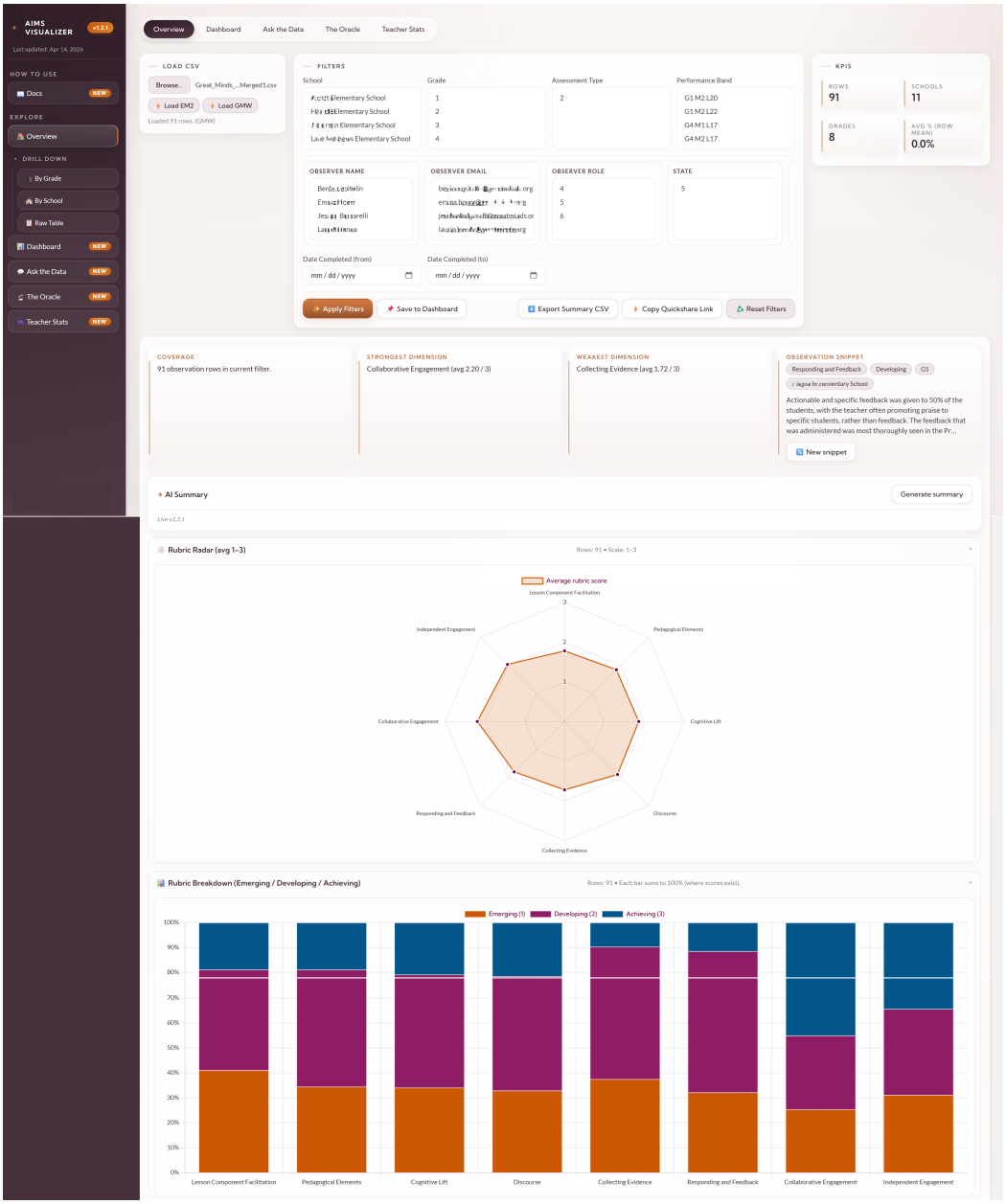
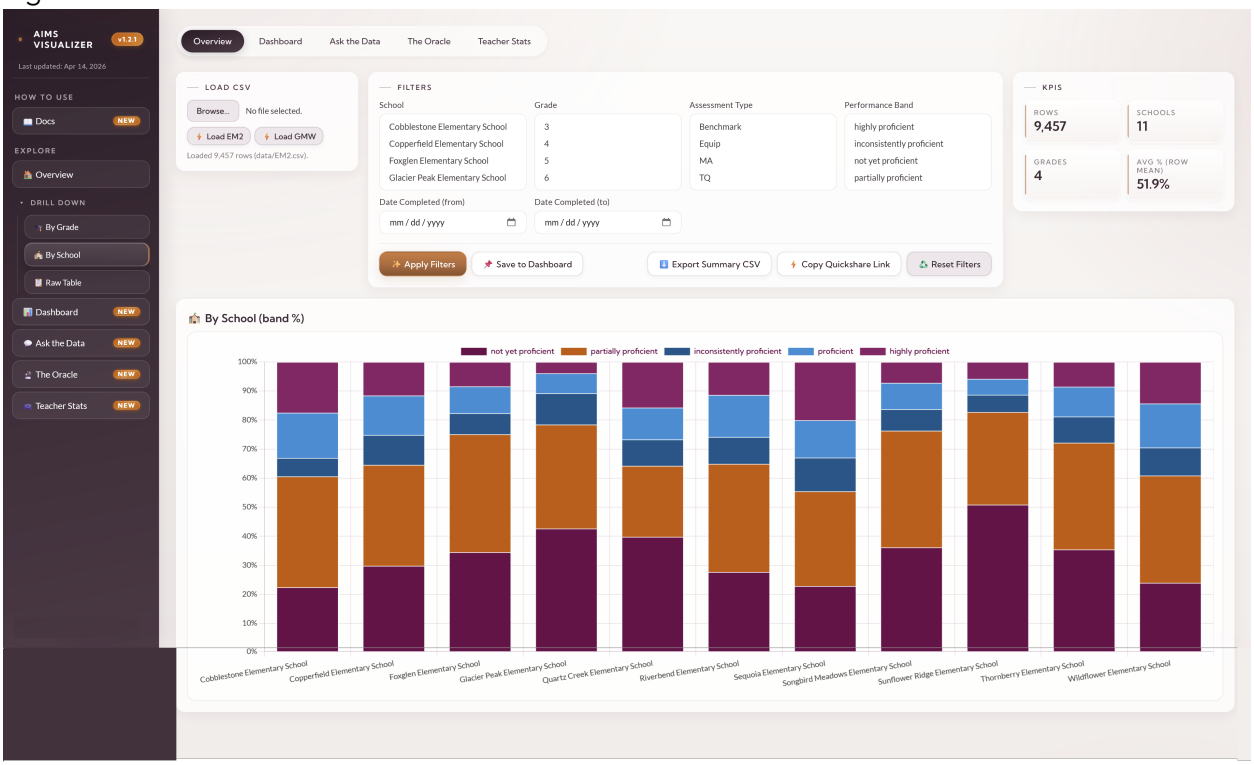


Figure B2: Main view of the Visualizer interface: assessment data.



From the main assessment and implementation pages, users can create charts using available filters to select a single or multiple schools and/or grade levels and save each chart to their dashboard. Users can then compare those saved school- and grade-level charts side-by-side.

Figure B3: Example side-by-side comparison of implementation/ walkthrough charts saved to the dashboard.

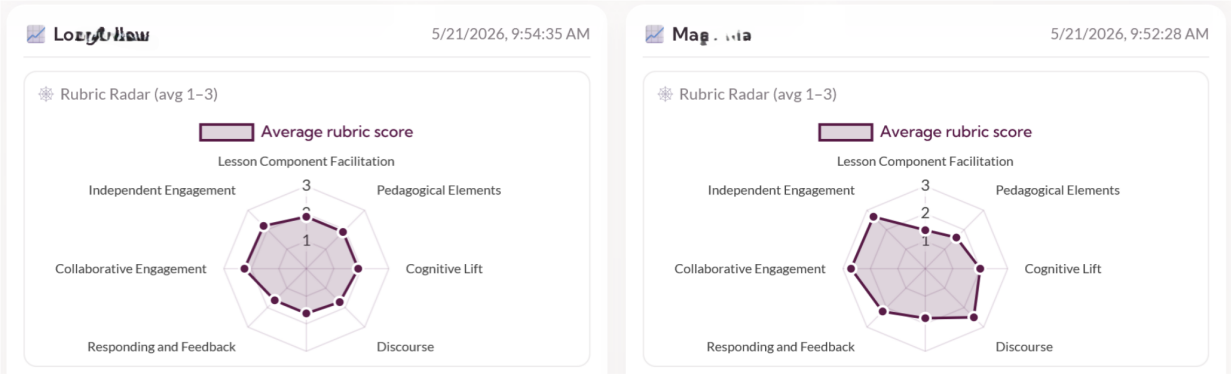
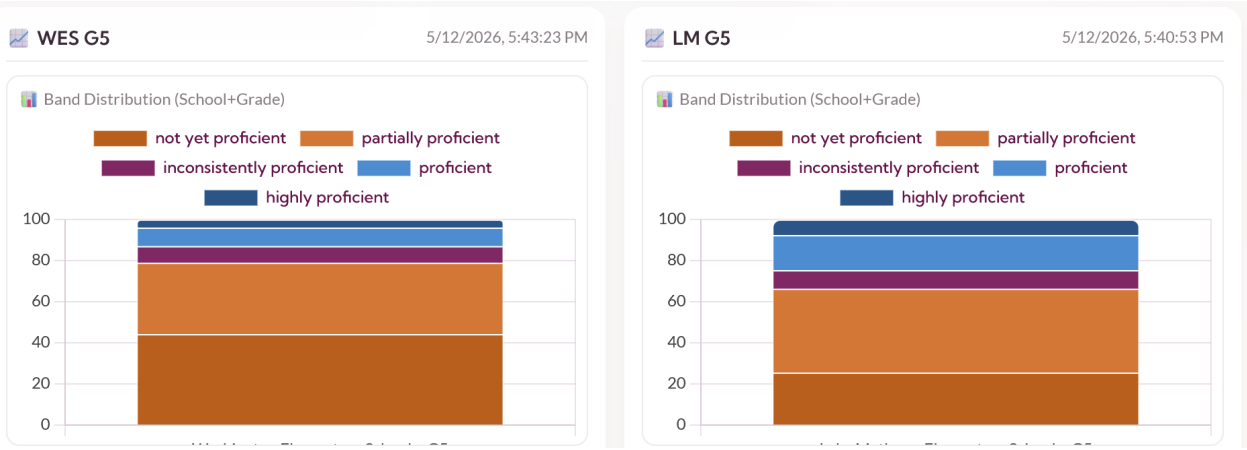
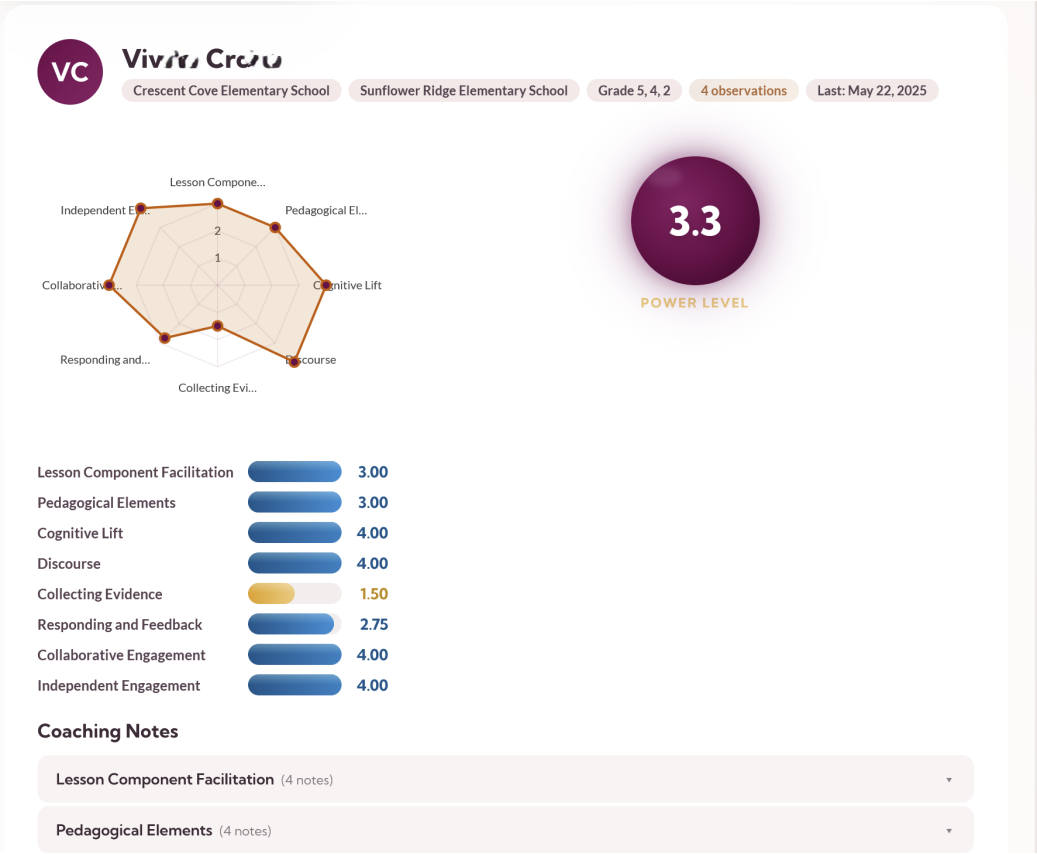


Figure B4: Example side-by-side comparison of assessment charts saved to the dashboard.



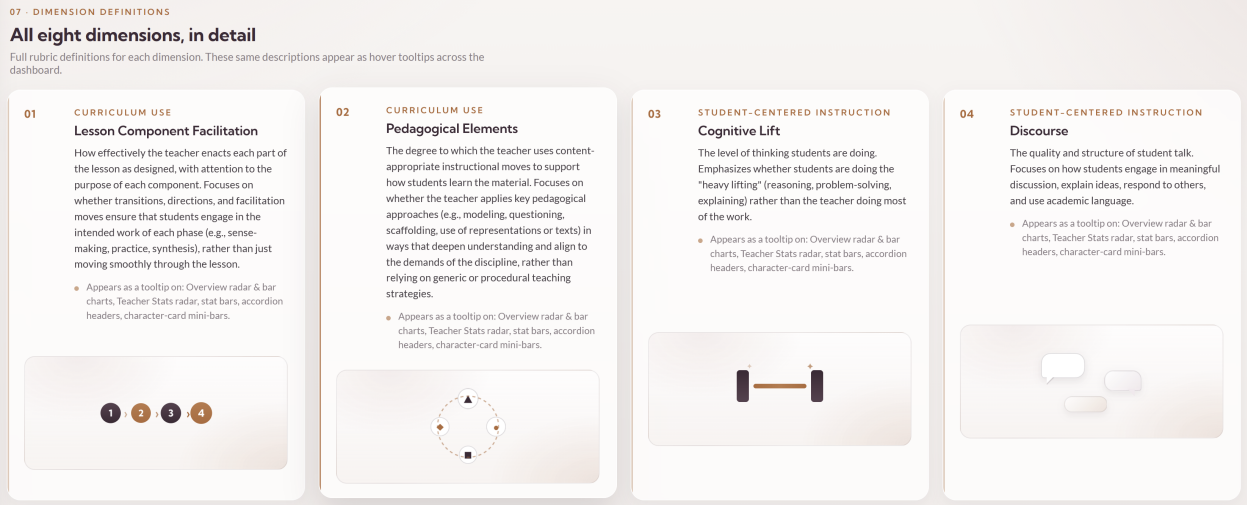
The “Teacher Stats” feature displays the performance profile of a selected teacher, including rubric statistics, observation history, and coaching notes.

Figure B5. “Teacher Stats” features individual teacher performance profiles.



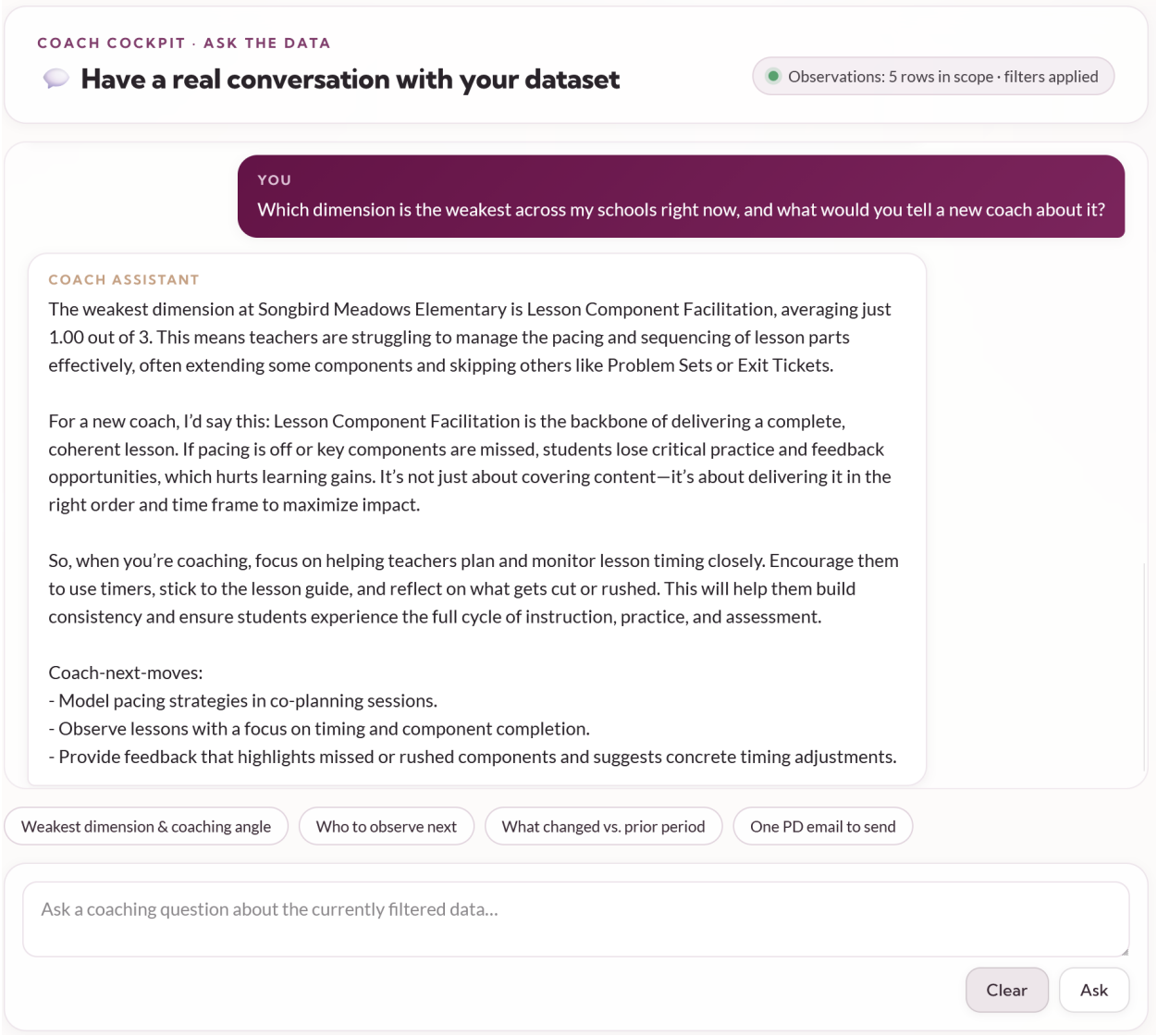
The terms used in the walk-through protocol are not necessarily familiar to all teachers and education leaders. To address this challenge, the definitions of each construct are accessible through hovering the computer pointer on the terms within the implementation page. They are also available in the documentation section of the Visualizer Platform.

Figure B6: Walk-through/implementation term definitions.



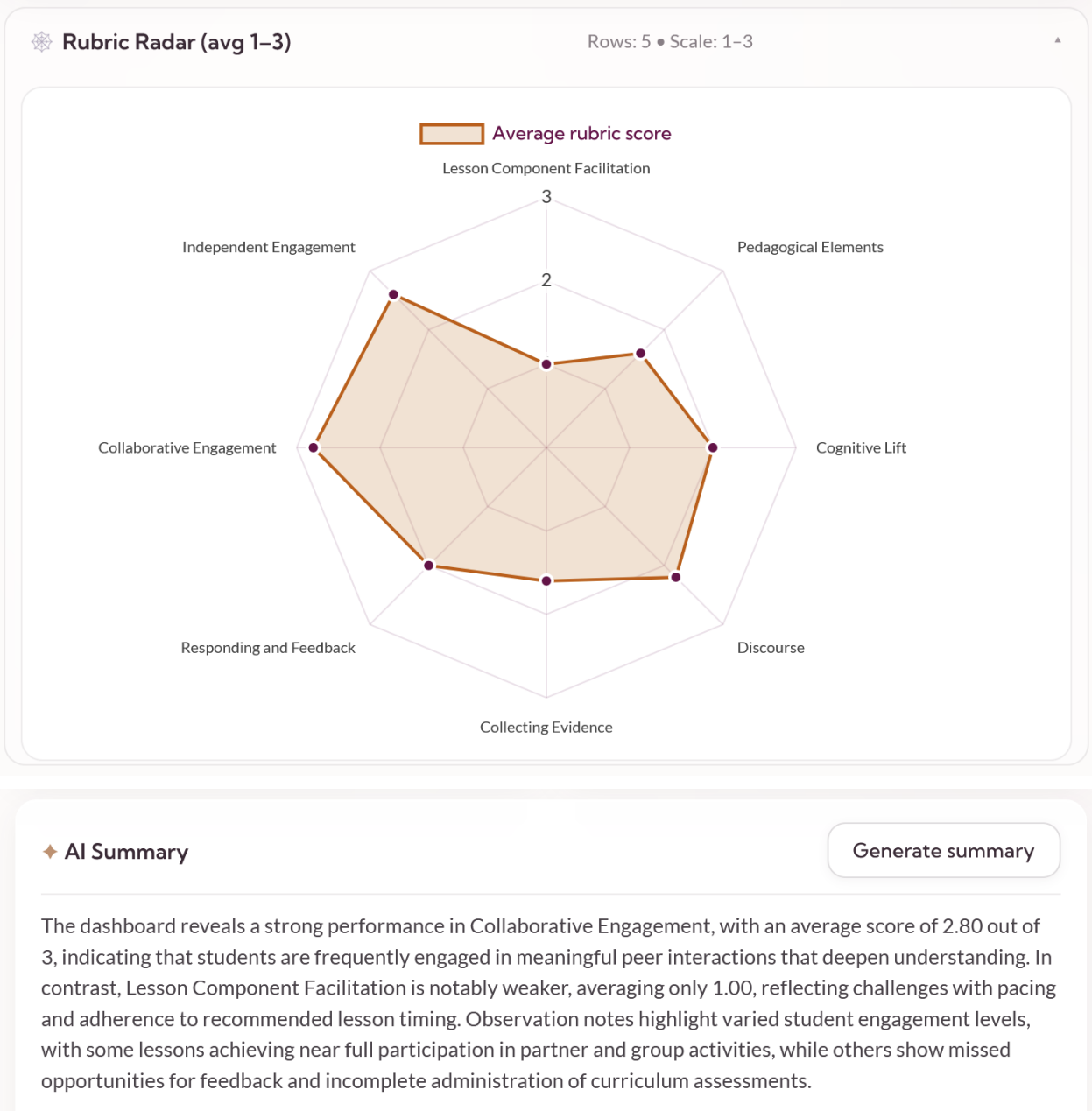
The Visualizer's "Ask the Data" feature allows users to engage in an LLM-driven conversation about the underlying data. The discussion is tailored to the current filter context, so users can ask questions and receive responses based on the specific subset of data they are viewing. This feature can summarize implementation/walk through and assessment data.

Figure B7. LLM-driven, filter-aware, "Ask the Data" feature.



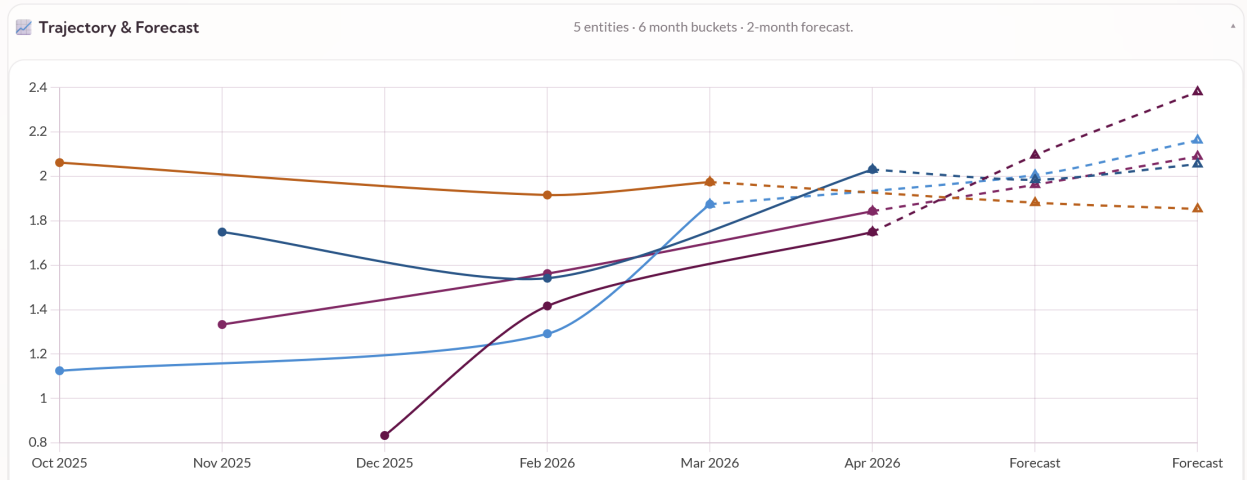
The platform includes an AI summary tool that can generate key takeaways from overall or filtered implementation/walkthrough data. This feature helps users identify patterns in complex charts and reduces the cognitive load required to interpret the results. This feature can summarize implementation/walk through and assessment data.

Figure B8. "AI Summary Tool" identifies key implementation data takeaways.



The “Oracle” feature displays trend-based forecasts generated from observations collected to date. The oracle projects forward each entity’s (e.g., school, grade-level, etc.) history using a simple linear regression, making the forecast most reliable for entities with at least four data points and a stable data-collection cadence. This feature can summarize implementation/walk through and assessment data.

Figure B9. Trend-based “Oracle Feature” linear regression forecasting.



Finally, the Visualizer platform includes documentation explaining each of its features.

Figure B9. Visualizer documentation and support.