

Pearl Education Launches GRADE, an Open Benchmark for Evaluating AI on Education Program Data

New benchmark measures whether AI can produce trustworthy, evidence-based insights from education program data

RICHMOND, VA, UNITED STATES, June 25, 2026 /EINPresswire.com/ -- [Pearl Education](#), today announced the launch of [GRADE](#) (Grounded Reasoning & Analysis for Data in Education), a new open-source benchmark designed to evaluate how effectively AI models analyze education program data and generate insights educators can trust and act on.



While AI benchmarks have become a standard way to evaluate large language models, most focus on general capabilities such as coding, mathematics, or academic reasoning. They do not answer a more practical question facing schools and education organizations every day:

Can an AI system accurately analyze education program data and produce conclusions that educators can trust and use to make decisions?

GRADE was created to answer that question.

"Many AI benchmarks tell us whether a model can solve a complex problem or write functional code," said Chris Kerr, Vice President of Product at Pearl Education. "GRADE measures something different. It evaluates whether an AI system can accurately analyze education program data, support its conclusions with evidence, acknowledge uncertainty when appropriate, and provide insights that would be useful to district analysts, program managers, and education leaders."

Why Pearl Education Built GRADE

Pearl Education works with districts, states, and education organizations to coordinate, monitor, and improve student support programs. Every day, education leaders ask questions such as:

- Which students are receiving support?
- Are programs being implemented as intended?
- Which interventions are producing results?
- Where should resources be adjusted to improve outcomes?

As AI capabilities advanced, Pearl Education saw an opportunity to help education leaders answer those questions more quickly and effectively. But the company also encountered a challenge shared by many organizations evaluating AI today.

Most benchmarks could measure whether a model could write code, solve math problems, or answer general knowledge questions. Few could determine whether an AI system could accurately analyze education program data and provide insights educators could confidently use in practice.

GRADE was created to fill that gap.

By measuring factual accuracy, evidence-based reasoning, analytical insight, consistency, and appropriate acknowledgment of uncertainty, GRADE evaluates the capabilities that matter most when AI is used to support real education decisions. Combined with the Arena's community-driven evaluation of usefulness, communication quality, and practitioner preference, GRADE provides a more complete view of how AI performs in real educational contexts.

For Pearl Education, the benchmark serves as both an internal standard for AI development and a contribution to the broader education community. By open-sourcing GRADE, the company hopes to encourage more rigorous, transparent evaluation of AI systems intended for use in education.

Measuring What Matters for Education

GRADE evaluates AI systems using realistic education program scenarios built around structured program data, including attendance records, session information, subgroup summaries, and published research references.

The benchmark includes 26 questions spanning five categories of analysis, ranging from straightforward fact retrieval to equity interpretation, program evaluation, and research synthesis.

All benchmark data is fully synthetic and seeded. No real students, schools, tutors, districts, or education programs are included.

Unlike traditional AI benchmarks that primarily focus on accuracy, GRADE evaluates qualities that education leaders consistently identify as essential for trustworthy decision-making:

- Factual Accuracy
- Evidence-Based Reasoning
- Analytical Insight
- Appropriate Acknowledgment of Uncertainty
- Consistency

The benchmark's automated scoring framework weights six dimensions:

- Accuracy (35%)
- Insight (20%)
- Evidence (15%)
- Honesty About Limits (15%)
- Consistency (10%)
- Clarity (5%)

Open by Design

Pearl Education designed GRADE as an open benchmark from the beginning.

Question definitions, scoring rubrics, methodology documentation, synthetic datasets, and evaluation criteria are publicly available. Any AI system can be evaluated using the benchmark, and results can be submitted to the public leaderboard through a standardized evaluation process.

For more subjective interpretation and recommendation tasks, GRADE also includes an Arena component. Education practitioners compare anonymized AI responses to the same scenario and vote on which answer is more useful.

While GRADE's automated scoring framework measures factual accuracy and analytical quality, the Arena captures a different dimension: what educators actually find most useful in practice. Together, the two approaches evaluate both the quality of an AI's analysis and how effectively it communicates insights to education leaders.

Advancing Trustworthy AI in Education

As schools increasingly explore AI to support data analysis, program evaluation, and operational decision-making, Pearl Education believes benchmarks should measure more than raw intelligence.

They should measure trustworthiness.

Education leaders are not asking whether AI can pass a test. They are asking whether AI can help

them make better decisions.

Pearl Education created GRADE because it believes the future of AI in education will be determined not by how impressive AI appears in a demonstration, but by how reliably it helps educators understand what is happening and what to do next.

The company is releasing GRADE as an open benchmark to help advance a shared standard for trustworthy, evidence-based AI across the education sector. The full benchmark, methodology, datasets, scoring framework, and evaluation tools are available at: <https://gradebench.ai>.

About Pearl Education

Pearl Education works with states, school districts, and tutoring providers to design, manage, and sustain student support programs at scale, including tutoring, intervention, and supplemental instruction.

The U.S.-based education technology company serves as the trusted technology infrastructure behind many of the nation's largest student support initiatives. Through its Student Support Platform, Pearl helps education leaders coordinate program delivery, measure participation and outcomes, and gain clearer insight into what's working for students. By streamlining scheduling and attendance, while centralizing and standardizing program data across schools, programs, and providers, Pearl helps districts and states strengthen implementation, support accountability, and inform funding decisions.

Trusted by schools and states nationwide, Pearl supports close to 1,000 schools and has powered more than 100 million minutes of student learning. Recognized by TIME as one of the Top EdTech Companies of 2025, Pearl partners with many of the nation's largest student support initiatives to help districts deliver effective, equitable, and sustainable programs.

Erin Grubbs

Pearl Education

erin@poweredbypearl.com

Visit us on social media:

[LinkedIn](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/921958846>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.