

Asset-oriented AI feedback for Black & Latino students at scale: Results of user testing

Problem

Researchers have described the need for Black and Latino students to receive asset-oriented feedback in their learning, feedback that is rich in information and that communicates high expectations for student achievement. Desmos Math contains an interface that lets teachers write feedback to students through its software platform. Much of that feedback doesn't meet the criteria described by those researchers. So we created an activity that would a) help teachers learn how to give asset-based mathematics feedback, b) introduce educators to the affordances of AI in their practice.

Procedures

We invited middle school mathematics teachers to participate in this teacher feedback activity through one of our email communications. Then, between April 8 and April 15, we invited those teachers to access the activity and complete a survey about their experience. We invited 95 educators. 44 teachers participated and took the survey. Demographic details about the teachers include:

- They taught in 19 different states.
- Their years of experience ranged from “1-5” to “more than 41” with “6-15” as the mode response (40% of teachers).
- The majority identified as female.
- The majority taught in public schools.
- They taught in a mix of rural (18%), suburban (46%), urban (36%) schools.
- 95% of the teachers said they agreed with the statement: “Before this experience, I felt confident I could give students effective feedback on their learning.”
- 85% of the teachers themselves identified as “White.”
- 58% of teachers categorized their schools as “predominantly White.”

Responsive to the grant’s RFP, we wanted to learn specifically from teachers of predominantly Black/Hispanic student populations and sites with a mix of racial demographics (“Target Teachers”) contrasted with predominantly white institutions (“Not Target Teachers”).

Results

We presented teachers with the most common incorrect answer to a question in our curriculum: Students commonly answer “4-7x” when asked to simplify “4 - 2x + 5x.”

Combine like terms and write an equivalent expression with the fewest number of terms.

$$4 - 2x + 5x$$

We then asked teachers to write feedback to those students and submit it, at which point we prompted an LLM with a) the text of the math problem, b) the student’s response, and c) the teacher’s feedback, each dynamically populated. We also included in our prompt d) a description of principles of asset-oriented feedback.

Here are several instances of teacher feedback to the incorrect student along with feedback from the LLM on the teacher’s feedback.

Teacher Feedback	AI Feedback
OOOOH. You made a GREAT mistake! Can you identify your mistake?	I appreciate your enthusiasm in your feedback, as it shows you are invested in the student's learning. However, remember our goal is to communicate high standards while also being specific and substantive. In your feedback, you've pointed out that there's a mistake, but it might be more helpful to guide the student towards finding and understanding that mistake. For example, you might ask, "Can you

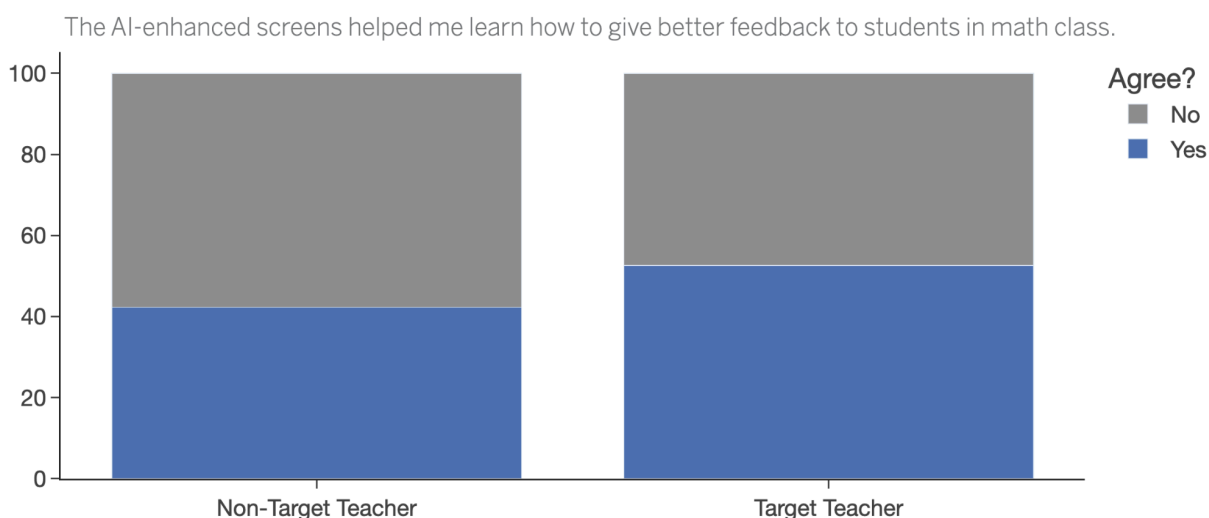
	<p>tell me how you combined the terms involving 'x'?" This question still prompts the student to find their mistake, but in a more specific and focused manner.</p>
<p>Not correct, give it another try.</p>	<p>I appreciate your dedication to guiding your students towards the correct answer. When pointing out that their answer was incorrect, it could be beneficial to also highlight what they did correctly. For instance, they correctly identified that they need to combine the 'x' terms. Instead of simply telling them to try again, perhaps ask them to revisit the steps of combining like terms. This way, your feedback is not only guiding them towards the correct answer, but also reinforcing their understanding of the concept.</p>
<p>How did you get that answer?</p>	<p>The feedback you provided shows that you are invested in the student's understanding, but it could be more constructive. Instead of just asking how they reached their answer, try to guide them in identifying where they might have made an error. For instance, you could ask, "Can you walk me through your process of combining like terms? Where might you have subtracted instead of added?" This encourages the student to revisit their process, emphasizes the mastery goal, and fosters their problem-solving skills.</p>
<p>Good start! You correctly identified that the x terms need to be combined. However, let's think about the sign of each term. How does $-2x$ plus $5x$ combine? Give it</p>	<p>Your feedback to the student demonstrates your belief in their ability to solve the problem, which is great. Additionally, you've pointed out what they did correctly</p>

another try. I am confident in your ability to solve this.	and where they need to improve. However, your feedback could be clearer by specifically pointing out the calculation error. Instead of asking, "How does $-2x$ plus $5x$ combine?", you might say, "Remember, when we combine $-2x$ with $5x$, we are actually adding $-2x$ and $5x$. Could this change your answer?" This way, you guide the student towards the correct answer while still encouraging them to solve it independently.
--	--

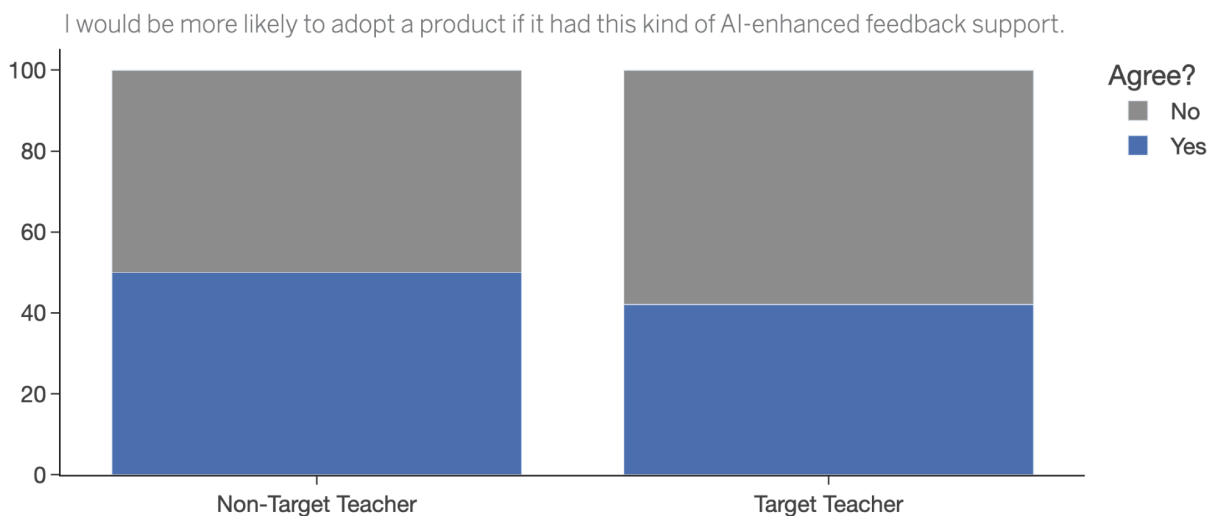
Table 1. A table displaying the kinds of feedback teachers wrote to students and the feedback the LLM gave teachers on their feedback.

We asked teachers about their agreement with the statement, “The AI-enhanced screens helped me learn how to give better feedback to students in math class.”

We used a five-point Likert-style item. (“Strongly Disagree” to “Strongly Agree” with an option for “Neutral.”) Slightly more than half of Target Teachers answered “Strongly Agree” or “Agree” to this statement.



We saw similar results when we asked teachers about their agreement with the statement, “I would be more likely to adopt a product if it had this kind of AI-enhanced feedback support” with agreement from 42% of Target Teachers and 50% of Non-Target Teachers.



Following are several responses from Target Teachers who **agreed** the training tool was helpful when we asked, “What was one aspect of this experience that was especially helpful?”

The AI response to my feedback was specific and mentioned details from my feedback.

After each of the practice feedback prompts I found myself trying to get a higher rating from the AI.

I like thinking of ways to give feedback to common wrong answers.

Following are several responses from Target Teachers who **disagreed** the training tool was helpful when we asked, “What is one way this experience could have been MORE helpful for you?”

I could easily frustrate myself by trying to get my feedback exactly right (at least according to the AI) when what I had already come up with is perfectly fine.

One way this experience could have been more helpful to me if I could have tracked my progress on my feedback ability or giving me a scale.

The AI reply to my feedback often suggested doing something I did in my response, so I had a “isn’t that what I just said?” type of response to the AI feedback.

Related to that last comment: 27% of teachers answered “Yes” when asked “Did you notice any errors (sometimes called “hallucinations”) on the AI-enhanced screens?”

Other recommendations for improvements

Given the response to the AI-enhanced screens (roughly 50% perceiving them as useful) we looked at other recommendations for improvement.

Several educators seemed to perceive the AI as a black box, with a lack of transparency about its criteria and judgment:

I would have liked to see an explanation of what the AI was going to do in a more descriptive way.

I didn't trust the feedback. For example, on one screen it told me "to use language the student can easily understand," but I used exact phrasing from our curriculum that our students would be used to. It also seemed to always look for a negative, even when my feedback was perfectly solid.

An educator noted the AI didn’t have all the necessary context to support their feedback.

It would help if the AI had information about where in the learning process the feedback is being given. I would give different feedback at the beginning of a unit than I would at the end.

Other educators noted that the AI never seemed satisfied.

I felt a little exhausted when every bit of feedback I gave was found to need improvement by the AI. I think there was one response on the quotient of powers problem that it did not understand my statement.

I tried lots of different feedback examples trying each suggestion from the AI helper. If I tried being more “leading,” it asked me to be

more instructive; if I tried adding hints, it asked me to guide the student to mastery, so I never got the AI to say I actually nailed it. Another example was that if I used "kid-friendly language" it asked me to be more rigorous, but if I used "math vocabulary" it said I was too abstract. Sometimes it suggested I acknowledge what the student did right when I actually did begin with that (almost like the AI missed that part).

It seemed like the chatbot was instructed to offer a critique (or additional information), no matter what the feedback submitted was. For example, on one of the pages, I entered some basic feedback, and the note I got back from the chatbot was that I might consider including more detail so the student really understands what to do next time. I edited my response and included some more detail, and the feedback from the chatbot was that while my detailed feedback was nice, I might consider reducing the amount of detail and keeping the feedback short and to the point.

In other instances, the AI would hallucinate and ask teachers to make changes that they had already made:

Also, there were times when it would suggest that I provide information in my feedback, that was provided. For example, I was told to redirect students back to laws of exponents when in my original feedback I put "Reconsider the laws of exponents when dividing powers with the same base."

The AI feedback didn't always understand what I was saying, so it told me to include things in my feedback that were already present.

Any surprises?

Finally, when asked what surprised them about the AI-enhanced training tool, educators teaching Black and Latino students said the following:

I was surprised by how readable and understandable the AI feedback was. I knew exactly what it was trying to say.

That there were so many possible pathways to provide valuable feedback.

When it quoted part of my feedback as a helpful idea - that it was that specific surprised me.

How positive it was with feedback to me.

Future development

Given the overall ambivalent response to this AI feedback training tool, we would like to investigate further the kinds of support educators need to provide effective feedback to their students. It may be that generative AI isn't the right medium for this kind of feedback, or it might be that our instantiation of generative AI wasn't right. Possible next steps:

- Add additional information into the LLM prompt to refine the feedback, for example the kinds of feedback the educator has previously authored and the type of lesson the teacher is teaching.
- Interview students who are receiving feedback from educators who have participated in the AI-enhanced training in writing feedback. Do they perceive a difference?
- Use more advanced generative AI techniques. For example, first ask the AI to categorize whether the teacher's feedback as asset-oriented, then give the teacher an indicator of success if it's appropriate, rather than suggest an improvement in every response.
- Use chain-of-thought prompting in a future iteration. We suspect our LLM prompt was too lengthy and complex.
- Trial this activity with novice teachers, because this particular group of teachers already felt confident in their ability to give feedback to students.

Overall, we have significant qualitative evidence in survey responses that teachers are *interested* in learning more about giving effective feedback, and that they are at least curious about generative AI. This particular instantiation has significant room for improvement, and we will continue to investigate ways to improve the kinds of feedback Black and Latino students receive from their teachers.