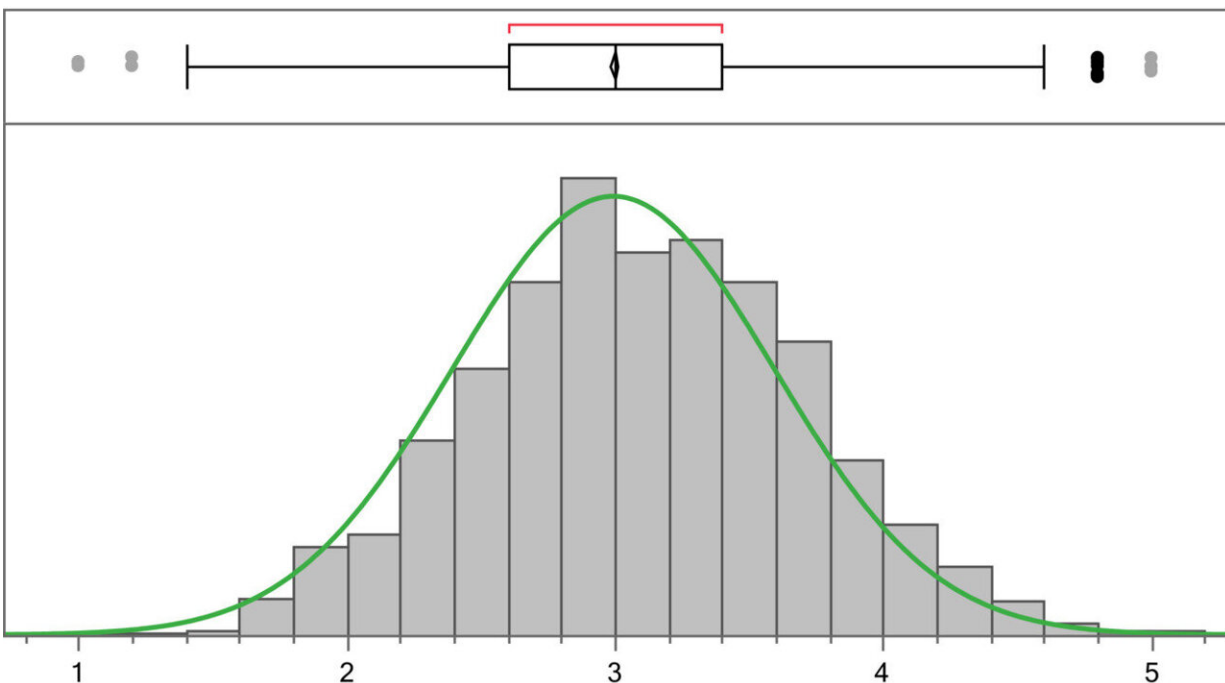


AI could improve assessments of childhood creativity

July 6 2023, by Erica Techo



Histogram of the Originality Ratings. *Note.* Data displayed for 10,449 responses, averaged across the five judges. The originality ratings were highly normally distributed, as indicated by the superimposed continuous normal distribution. Credit: *The Journal of Creative Behavior* (2023). DOI: 10.1002/jocb.588

A new study from the University of Georgia aims to improve how we evaluate children's creativity through human ratings and through artificial intelligence.

A team from the Mary Frances Early College of Education is developing an AI system that can more accurately rate open-ended responses on [creativity](#) assessments for elementary-aged students.

"In the same way that hospital systems need good data on their patients, educational systems need really good data on their students in order to make effective choices," said study author and associate professor of educational psychology Denis Dumas. "Creativity assessments have policy and curricular relevance, and without assessment data, we can't fully support creativity in schools."

These tests are commonly used to identify gifted or talented students who require additional instructional resources to be served adequately by schools. And because they are time-intensive to evaluate—most open-ended responses require grading from multiple trained human judges—they are not as widely used as their math, reading or IQ counterparts. By creating an AI system, however, creativity assessments could become a more accessible tool for schools.

To improve the AI's functionality, Dumas and his collaborators analyzed more than 10,000 individual responses on a 30-minute creativity assessment. They found that some categories of students and some types of responses led to less consistent creativity ratings among judges. All identifiable [student](#) information was removed from the assessments, and judges only received student responses.

"Our judges didn't know who the kids were and did not know their specific demographics," Dumas said. "There wasn't an explicit bias, but something about the way some students responded made their responses harder for our team to rate reliably."

Judges were instructed to score responses between 1 (most unoriginal) and 5 (most original), and they were more likely to disagree on ratings

when responses showed less originality or those that came from [younger children](#) or [male students](#).

"I suspected there would be more disagreement among graders at the top of the originality scale, but we found that because judges were looking for originality, they were more likely to agree when a response was unusual, surprising and clever," Dumas said. "But when an answer [scored] lower on the originality scale, that caused more disagreement."

For example, when asked for a surprising use for a hat, a third grader suggested "you cut off the shade part and it will look silly." Judgments on this response ranged from a 1 to a 4, and the study highlighted this example of how younger students' responses can be more difficult to rate. Some judges viewed this as unoriginal, as the hat remains a wearable item to put on your head. Others, however, saw the alteration of a hat's appearance as funny, surprising and age-appropriate for a creative third grader.

A wider range of scores also appeared with highly original responses from [gifted students](#), with LatinX students identified as English Language Learners and with Asian students who took more time on the tasks. All of these factors led to more ratings disagreement.

"Children who are bilingual, they are going to write their responses differently; their responses are formulated differently than a child who is monolingual," Dumas said. "Even though many of our readers were also bilingual, that can be hard to apply in the ratings context. It seemed like what we were finding over and over again is that the students who were more likely to be bilingual were also harder to rate."

Understanding where ratings disagreement cropped up helps retrain the AI system and make it more accurate, Dumas said, which helps reduce the error band on assessment results. These error bands are standard fare

on assessments commonly used in schools, Dumas said, but can be wider on creativity assessments than, say, math or reading tests. The narrower the band, the more confident schools can be when making decisions based on the scores.

This study is one step toward improving accuracy, and thereby confidence, in these assessments, Dumas said.

"What gets assessed in schools tends to be what teachers focus on in their instruction. So the values and priorities of a school system can be observed in the assessments they choose," Dumas said. "I would love to be able to build a creativity [assessment](#) more into the [school](#) psychologist toolkit and give them an option to observe creative potential in a young child and interpret that as a strength."

The study is published in *The Journal of Creative Behavior*.

More information: Denis Dumas et al, What Makes Children's Responses to Creativity Assessments Difficult to Judge Reliably?, *The Journal of Creative Behavior* (2023). [DOI: 10.1002/jocb.588](https://doi.org/10.1002/jocb.588)

Provided by University of Georgia

Citation: AI could improve assessments of childhood creativity (2023, July 6) retrieved 27 April 2024 from <https://phys.org/news/2023-07-ai-childhood-creativity.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.