

1086-J5-577

**James H Fife\*** ([jfife@ets.org](mailto:jfife@ets.org)), Educational Testing Service, Princeton, NJ 08541. *The automated scoring of constructed-response mathematics test questions: Why humans sometimes disagree with the computer.*

The automated scoring of constructed-response mathematics test items is a relatively mature capability that can score numeric responses and equations with 100% accuracy. But does an automated scoring engine always produce the same score that human scorers would? To find out, approximately 300 responses to each of 97 constructed-response mathematics items were double human scored and were scored by an automated scoring engine. The human and automated scores were then compared to determine how frequently and why human scores disagreed with automated scores. Of the 97 items, there were 27 items for which the human scores agreed with each other but disagreed with the automated scores. My research focused mainly on these items. How often do humans agree with each other but disagree with the automated score? Why do the humans disagree with the automated score? Based on my analysis, I developed a taxonomy of human/automated disagreement, which I will discuss in this presentation. (Received September 07, 2012)