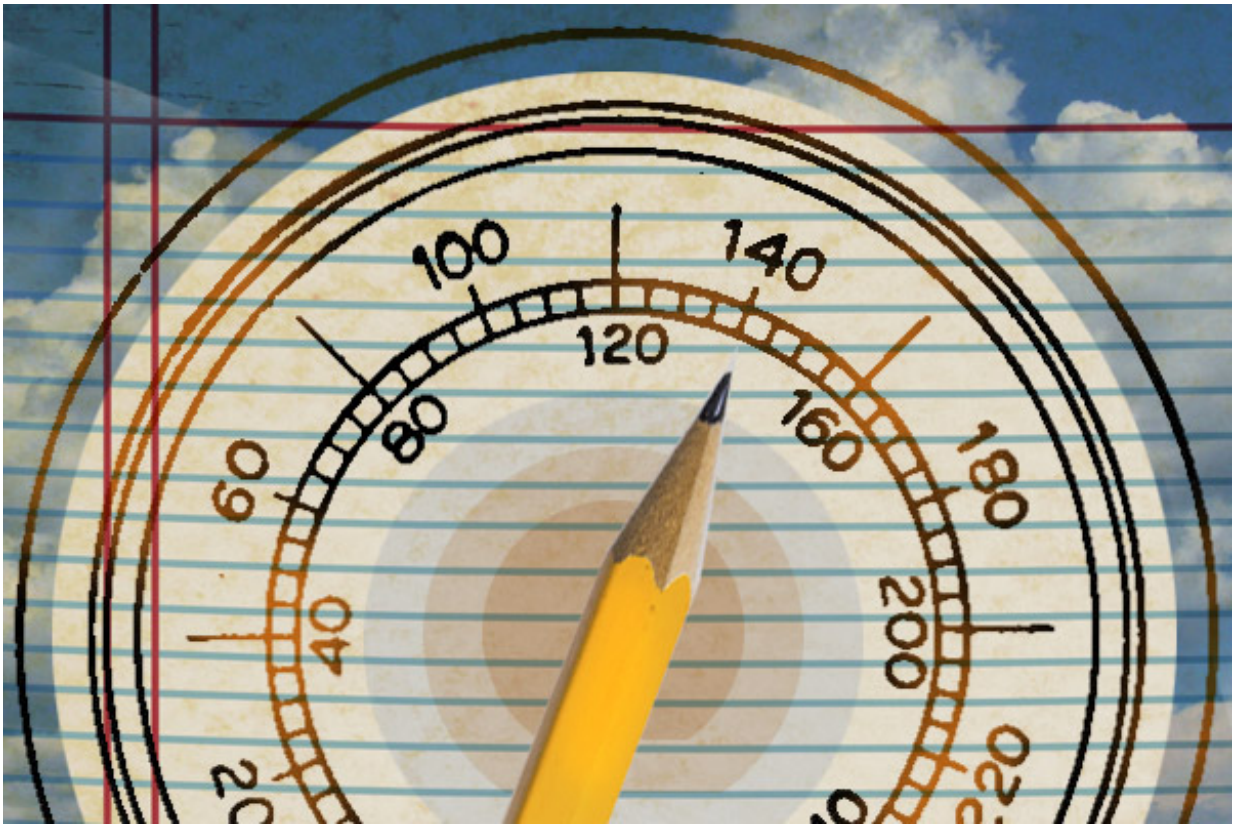# Researcher studies teachers' use of automated essay scoring software

July 8 2015, by Beth Miller

Joshua Wilson, assistant professor in the University of Delaware's School of Education, has been studying the use of algorithm-based software that evaluates writing.

"Can we write during recess?" Some students were asking that question

at Anna P. Mote Elementary School, where teachers were testing software that automatically evaluates essays for University of Delaware researcher Joshua Wilson.

Wilson, assistant professor in UD's School of Education in the College of Education and Human Development, asked teachers at Mote and Heritage Elementary School, both in Delaware's Red Clay Consolidated School District, to use the software during the 2014-15 school year and give him their reaction.

Wilson, whose doctorate is in special education, is studying how the use of such software might shape instruction and help struggling writers.

The software Wilson used is called [PEGWriting](#) (which stands for Project Essay Grade Writing), based on work by the late education researcher Ellis B. Page and sold by Measurement Incorporated, which supports Wilson's research with indirect funding to the University.

The software uses algorithms to measure more than 500 text-level variables to yield scores and [feedback](#) regarding the following characteristics of writing quality: idea development, organization, style, word choice, sentence structure, and writing conventions such as spelling and grammar.

The idea is to give teachers useful diagnostic information on each writer and give them more time to address problems and assist [students](#) with things no machine can comprehend – content, reasoning and, especially, the young writer at work.

Writing is recognized as a critical skill in business, education and many other layers of social engagement. Finding reliable, efficient ways to assess writing is of increasing interest nationally as standardized tests add writing components and move to computer-based formats.

[The National Assessment of Educational Progress](link), also called the Nation's Report Card, first offered computer-based writing tests in 2011 for grades 8 and 12 with a plan to add grade 4 tests in 2017. That test uses trained readers for all scoring.

Other standardized tests also include writing components, such as the assessments developed by the Partnership for Assessment of College and Careers (PARCC) and the Smarter Balanced Assessment, used for the first time in Delaware this year. Both PARCC and Smarter Balanced are computer-based tests that will use automated essay scoring in the coming years.

Researchers have established that computer models are highly predictive of how humans would have scored a given piece of writing, Wilson said, and efforts to increase that accuracy continue.

However, Wilson's research is the first to look at how the software might be used in conjunction with instruction and not as a standalone scoring/feedback machine.

In earlier research, Wilson and his collaborators showed that teachers using the automated system spent more time giving feedback on higher-level writing skills – ideas, organization, word choice.

Those who used standard feedback methods without automated scoring said they spent more time discussing spelling, punctuation, capitalization and grammar.

The benefits of automation are great, from an administrative point of view. If computer models provide acceptable evaluations and speedy feedback, they reduce the amount of needed training for human scorers and, of course, the time necessary to do the scoring.

Consider the thousands of standardized tests now available – state writing tests, SAT and ACT tests for college admission, GREs for graduate school applicants, LSATs for law school hopefuls and MCATs for those applying to medical school.

When scored by humans, essays are evaluated by groups of readers that might include retired teachers, journalists and others trained to apply specific rubrics (expectations) as they analyze writing.

Their scores are calibrated and analyzed for subjectivity and, in large-scale assessments, the process can take a month or more. Classroom teachers can evaluate writing in less time, of course, but it still can take weeks, as any English teacher with five or six sections of classes can attest.

"Writing is very time and labor and cost intensive to score at any type of scale," Wilson said.

Those who have participated in the traditional method of scoring standardized tests know that it takes a toll on the human assessor, too.

Where it might take a human reader five minutes to attach a holistic score to a piece of writing, the automated system can process thousands at a time, producing a score within a matter of seconds, Wilson said.

"If it takes a couple weeks to get back to the student they don't care about it anymore," he said. "Or there is no time to do anything about it. The software vastly accelerates the feedback loop."

But computers are illiterate. They have zero comprehension. The scores they attach to writing are based on mathematical equations that assign or deduct value according to the programmer's instructions.

They do not grade on a curve. They do not understand how far Johnny has come in his writing and they have no special patience for someone who is just learning English.

These computer deficiencies are among the reasons many teachers – including the National Council of Teachers of English – roundly reject computerized scoring programs. They fear a steep decline in instruction, discouraging messages the soulless judge will send to students, and some see a real threat to those who teach English.

In a recent study, Wilson and other collaborators showed that use of automated feedback produced some efficiencies for teachers, faster feedback for students, and moderate increases in student persistence.

This time they brought a different question to their review. Could automated scoring and feedback produce benefits throughout the school year, shaping instruction and providing incentives and feedback for struggling writers, beyond simply delivering speedy scores?

"If we use the system throughout the year, can we start to improve the learning?" Wilson said. "Can we change the trajectory of kids who would otherwise fail, drop out or give up?"

To find out, he distributed free software subscriptions provided by Measurement Incorporated to teachers of third-, fourth- and fifth-graders at Mote and Heritage and asked them to try it during the 2014-15 school year.

Teachers don't dismiss the idea of automation, he said. Calculators and other electronic devices are routinely used by educators.

"Do math teachers rue the day students didn't do all computations on their own?" he said.

Wilson heard mixed reviews about use of the software in the classroom when he met with teachers at Mote in early June.

Teachers said students liked the "game" aspects of the automated writing environment and that seemed to increase their motivation to write quite a bit. Because they got immediate scores on their writing, many worked to raise their scores by correcting errors and revising their work over and over.

"There was an 'aha!' moment," one teacher said. "Students said, 'I added details and my score went up.' They figured that out."

And they wanted to keep going, shooting for higher scores.

"Many times during recess my students chose to do PEGWriting," one teacher said. "It was fun to see that."

That same quick score produced discouragement for other students, though, teachers said, when they received low scores and could not figure out how to raise them no matter how hard they worked. That demonstrates the importance of the teacher's role, Wilson said. The teacher helps the student interpret and apply the feedback.

Teachers said some students were discouraged when the software wouldn't accept their writing because of errors. Others figured out they could cut and paste material to get higher scores, without understanding that plagiarism is never acceptable. The teacher's role is essential to that instruction, too, Wilson said.

Teachers agreed that the software showed students the writing and editing process in ways they hadn't grasped before, but some weren't convinced that the computer-based evaluation would save them much time. They still needed to have individual conversations with each

student – some more than others.

"I don't think it's the answer," one teacher said, "but it is a tool we can use to help them."

How teachers can use such tools effectively to demonstrate and reinforce the principles and rules of writing is the focus of Wilson's research. He wants to know what kind of training teachers and students need to make the most of the software and what kind of efficiencies it offers teachers to help them do more of what they do best: teach.

Bradford Holstein, principal at Mote and a UD graduate who received a bachelor's degree in 1979 and a master's degree in 1984, welcomed the study and hopes it leads to stronger writing skills in students.

"The automated assessment really assists the teachers in providing valuable feedback for students in improving their writing," Holstein said.

Provided by University of Delaware