Project Essay Grade (PEG)

Current Usage and Research, 2014

Shayne Miel

February 7, 2014





MEASUREMENT INCORPORATED
EXCELLENCE IN ASSESSMENT™
www.measurementinc.com

EXCELLENCE IN ASSESSMENT

PEG: USAGE AND RESEARCH 2014

Measurement Incorporated (MI) has been at the forefront of scoring student writing since the early 1980s. MI pioneered many of the complex processes involved in cost-effectively handscoring student essays on a large scale—scoring writing assessments for numerous U.S. state departments of education, including Texas, Ohio, Michigan, Florida, and New Jersey. By the late 1990s, MI's expertise in handscoring had firmly established the company as the industry's premier writing assessment company.

By early 2000, MI had also established a collegial relationship with Dr. Ellis Batten Page of nearby Duke University. Page, regarded as the "father of automated essay scoring" from his pioneering work in the early 1960s, was the first to explore, document, and validate the computer-based assessment of written prose. His software was entering a new era as advances in microcomputer technology and the emergence of the World Wide Web were making automated essay scoring a practical possibility. In 2003, MI acquired Project Essay Grade™ (PEG™)[1] from Dr. Page and his associates. Eleven years later, MI has re-engineered, enhanced and extended the PEG system using the latest techniques and technologies in the field of computational linguistics, machine learning, and natural language processing.

With subsequent improvements in PEG and general advances in the reliability of machine scoring, artificial intelligence (AI) scoring has become a valuable, and in some cases, essential, tool in a variety of contexts. MI's AI scoring engine, PEG, is currently in use in summative and formative assessments, and we are studying its use in computer adaptive assessments. It is being used in pilot and field tests for the Smarter Balanced Assessment Consortium, which represents 31 U.S. states, where AI scoring will provide a necessary core element enabling the scoring of millions of student written responses. PEG, with an established track record in scoring essays for qualitative characteristics such as such as organization, support, word choice, and mechanics, has also performed well in studies of AI scoring for content and will be at the forefront of these national assessment developments.

Summative Assessments

Since 2009, the Utah State Office of Education has successfully used PEG as the sole scoring method on the statewide summative Direct Writing Assessment in Grades 5 and 8. Over the past four years, PEG has scored 344,000 student responses on Utah's six trait rubric. In addition, in 2013 PEG was used as the second reader on the Connecticut SBAC Aligned Practice Assessment (APA), providing scores for 90,000 student responses on Connecticut's three trait rubric.

In Spring 2013, PEG was selected as one of the AI engines to be deployed by the Smarter Balanced Assessment Consortium to provide automated scoring of items on the pilot and field tests of its next generation assessments. PEG scored 213,000 essay and short answer (ELA and

1 Project Essay Grade and PEG are trademarks of Measurement Incorporated.

Math) responses for the pilot test in Fall 2013, and will be scoring approximately 2.5 million responses for the field test in Fall 2014.

In 2012, the Hewlett Foundation sponsored two global competitions in automated scoring – the Automated Student Assessment Prize (ASAP), Phases 1 and 2. These competitions were the first of their kind and were intended to independently evaluate the state of the art in essay and short answer scoring. In both phases, PEG outperformed the competitors by achieving the highest level of agreement with respect to the human scores (Shermis & Hamner, 2013; Morgan, Shermis, Van Deventer, & Vander Ark, 2013). In addition to the ASAP results, there is a wealth of independent research that examines the validity and reliability of automated scoring, particularly as it relates to summative assessment, including a large body of work conducted by Dr. Page himself over nearly 40 years. This growing body of research demonstrates the viability of AI scoring in general and MI's leadership in particular.

Formative Assessments

PEG has also been used to provide tens of millions of scores to students in formative writing assessments, with over three million essays scored in the last year alone. In addition to providing real-time scores, PEG also adds value when used in a formative context by providing response-specific feedback to the students on the grammar and spelling errors found in their essays, as well as offering targeted instructional feedback on how to improve their writing skills. PEG is in widespread use as an AI scoring engine for formative writing practice websites, including Educational Records Bureau's *Writing Practice Program (WPP)*, Utah State Office of Education's *Utah Write*, Connecticut State Department of Education's *CBAS Write*, North Carolina's *NC Write*, Learning Express's *Learning Express Advantage*, Measurement Planet's *Writing* Planet, and MI's own *PEG Writing*.

A recent study, conducted at University of Connecticut's NAEG School of Education, has shown that students were able to use PEG's automated scoring and feedback to increase their essay scores with repeated revisions of an essay, with the highest growth shown in the first few revisions (Wilson, Olinghouse, & Andrada, in press). In related research, Wilson and Andrada were able to use multiple essay revisions with PEG's automated scoring and feedback to more accurately identify struggling writers, in comparison to a static first-draft assessment (Wilson & Andrada, 2013). Two-thirds of the students initially identified as at-risk, were able to move out of the at-risk classification, given 5 or more revisions with feedback. These results point to the ability of PEG to not only assess writers in a typical summative assessment, but also to be used as an assessment and intervention tool in the context of a formative system.

MI is also pleased to have been selected to participate in the third phase of ASAP research, the Classroom Trials, which should serve to further formative assessment research. The emphasis in the first two phases of ASAP was on evaluating the degree to which current high-stakes writing assessments might be scored through automated methods. The Classroom Trials phase, on the other hand, examines the role of automated scoring in helping students achieve higher levels of writing proficiency and assisting teachers in the design and development of effective individualized instructional strategies.

References

Morgan, J., Shermis, M. D., Van Deventer, L., & Vander Ark, T. (2013) *Automated Student Assessment Prize: Phase 1 & Phase 2.* Retrieved from http://gettingsmart.com/wp-content/uploads/2013/02/ASAP-Case-Study-FINAL.pdf

Shermis, M. D., & Hamner, B. (2013). Contrasting State-of-the-Art Automated Scoring of Essays. *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, 313.

Wilson, J., & Andrada, G. (2013). *Examining Patterns of Writing Performance of Struggling Writers on a Statewide Classroom Benchmark Writing Assessment: The Utility of Dynamic Assessment.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wilson, J., Olinghouse, N. G., & Andrada, G. (in press). Does automated feedback improve writing quality? *Learning Disabilities: A Contemporary Journal.*