OWEAC is using work from Humanreaders.org as the basis for OWEAC's position against machine scoring of writing. The Humanreaders work was composed by well known and respected scholars in rhetoric and composition in consultation with national experts on writing assessment. It not only encapsulates OWEAC's position very clearly, it also includes the research to back up the incredibly well-reasoned position. Once we know the range of options available as assessment instruments BEYOND machine scoring, then OWEAC can weigh in with more specific recommendations to the Oregon State Department of Education.

RESEARCH FINDINGS ABOUT MACHINE SCORING OF WRITING SHOWS:

THAT no one—students, parents, teachers, employers, administrators, legislators—can rely on machine scoring of essays:

1. computer algorithms cannot recognize the most important qualities of good writing, such as truthfulness, tone, complex organization, logical thinking, or ideas new and germane to the topic (Byrne, Tang, Truduc, & Tang, 2010)

2. to measure important writing skills, machines use algorithms that are so reductive as to be absurd: *sophistication of vocabulary* is reduced to the average length or relative infrequency of words, or *development of ideas* is reduced to average sentences per paragraph (Perelman, 2012b; Quinlan, Higgins, & Wolff, 2009)

3. machines over-emphasize grammatical and stylistic errors (Cheville, 2004) yet miss or misidentify such errors at intolerable rates (Herrington & Moran, 2012)

4. machines cannot score writing tasks long and complex enough to represent levels of writing proficiency or performance acceptable in school, college, or the workplace (Bennett, 2006; Condon, 2013; McCurry, 2010; Perelman, 2012a)

5. machines require artificial essays finished within very short time frames (20-45 minutes) on topics of which student writers have no prior knowledge (Bridgeman, Trapani, & Yigal, 2012; Cindy, 2007; Jones, 2006; Perelman, 2012b; Streeter, Psotka, Laham, & MacCuish, 2002; Wang, & Brown, 2008; Wohlpart, Lindsey, & Rademacher, 2008)

6. in these short trivial essays, mere length becomes a major determinant of score by both human and machine graders (Chodorow & Burstein, 2004; Perelman, 2012b)

7. machines are not able to approximate human scores for essays that do fit real-world writing conditions; instead, machines fail badly in rating essays written in these situations (Bridgeman, Trapani, & Yigal, 2012; Cindy, 2007; Condon, 2013; Elliot, Deess, Rudniy, & Joshi, 2012; Jones, 2006; Perelman, 2012b; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002; Streeter, Psotka, Laham, & MacCuish, 2002; Wang & Brown, 2008; Wohlpart, Lindsey, & Rademacher, 2008)

8. high correlations between human scores and machine scores reported by testing firms are achieved, in part, when the testing firms train the humans to read like the machine, for instance, by directing the humans to disregard the truth or accuracy of assertions (Perelman, 2012b), and by requiring both machines and humans to use scoring scales of extreme simplicity

9. machine scoring shows a bias against second-language writers (Chen & Cheng, 2008) and minority writers such as Hispanics and African Americans (Elliot, Deess, Rudniy, & Joshi., 2012]

10. for all these reasons, machine scores predict future academic success abysmally (Mattern & Packman, 2009; Matzen & Hoyt, 2004; Ramineni & Williamson, 2013)

AND THAT machine scoring does not measure, and therefore does not promote, authentic acts of writing:

1. students are subjected to a high-stakes response to their writing by a device that, in fact, cannot read, as even testing firms admit (Elliott, 2011)

2. in machine-scored testing, often students falsely assume that their writing samples will be read by humans with a human's insightful understanding (Herrington & Moran, 2006)

3. conversely, students who knowingly write for a machine are placed in a bind since they cannot know what qualities of writing the machine will react to positively or negatively, the specific algorithms being closely guarded secrets of the testing firms (Frank, 1992; Rubin & O'Looney, 1990)—a bind made worse when their essay will be rated by both a human and a machine

4. students who know that they are writing only for a machine may be tempted to turn their writing into a game, trying to fool the machine into producing a higher score, which is easily done (McGee, 2006; Powers, Burstein, Chodorow, Fowles, & Kukich, 2001; see item 6, above)

5. teachers are coerced into teaching the writing traits that they know the machine will count— surface traits such as essay length, sentence length, trivial grammatical mistakes, mechanics, and topic-related vocabulary—and into *not*teaching the major traits of successful writing—elements such as accuracy, reasoning, organization, critical and creative thinking, and engagement with current knowledge (Council, 2012; Deane, 2013; Herrington & Moran, 2001; National, 2010)

6. machines also cannot measure authentic audience awareness, a skill essential at all stages of the composing process and correlative with writing competence of students both in the schools (Wolmann-Bonilla, 2000) and in college (Rafoth, 1985)

7. as a result, the machine grading of high-stakes writing assessments seriously degrades instruction in writing (Perelman, 2012a), since teachers have strong incentives to train students in the writing of long verbose prose, the memorization of lists of lengthy and rarely used words, the fabrication rather than the researching of supporting information, in short, to dumb down student writing.

You can find the statement, information about the composers of the document, a list of professionals with whom they consulted, and the above noted research at the following link: http://humanreaders.org/petition/research_findings.htm

To contact OWEAC:
Jillanne Michell, OWEAC Chair
Writing & Literature Instructor
Umpqua Community College
Roseburg, OR 97470
1-541-440-4646
jillanne.michell@umpqua.edu