



November 12, 2013

Contact: Mark Pudlow 850.201.3223 or 850.508.9756

FEA cautions against jumping to conclusions after court orders flawed VAM numbers released

TALLAHASSEE – Florida Education Association (FEA) President Andy Ford expressed disappointment after a First District Court of Appeal (DCA) panel ordered the release of flawed evaluation data for every teacher in Florida, overturning a trial judge’s earlier ruling. He cautioned Floridians not to jump to conclusions about the rankings of teachers because the numbers provided by the Florida Department of Education (DOE), based solely upon student test scores, provide an incorrect measure of public school teachers.

FEA intervened to join the Florida Department of Education in the case -- Morris Publishing Group v. Florida DOE -- defending against the disclosure of teachers’ value-added scores in response to a public records request by Jacksonville newspaper The Florida Times-Union. The trial court ruled against disclosure since teacher performance evaluations have historically been exempt from public disclosure, and The Times-Union appealed. Today, a panel of the First DCA overturned the trial court ruling.

“The evaluation data on teachers that is about to be made public is meaningless, which is why we joined in to enforce the public records exemption and prevent it from being published,” said FEA President Andy Ford. “The numbers to be released are subject to misinterpretation. They have not been put in their proper context.”

Ford said that research has shown that even the most sophisticated and valid “value-added” or “VAM” measurements are limited in what they can measure.

“But Florida’s VAM formula is not valid,” Ford said. “it is deeply flawed.”

Nearly all teachers’ VAM numbers are calculated according to students’ FCAT scores, yet only about 35 percent teach students and subjects tested on the FCAT. So for 65 percent or more of teachers, the VAM does not even attempt to measure the teacher’s actual teaching. The Legislature openly recognized this flaw earlier in passing SB 1664, which requires future VAM scores to be based upon a teacher’s actual students. But the two years of VAM data the court has ordered to be released does not take into account the new law, making all of the data meaningless.

“The FEA fully supports teacher accountability,” Ford said. “But assessments of teachers, like assessments of students, must be valid, transparent and multi-faceted. These value-added model calculations are none of these. We hope that The Florida Times-Union – and anyone else

who publishes these numbers – makes it fully clear to its readers how little meaning these numbers have in determining the quality of an individual teacher.”

FEA VAM Talking Points

- The evaluation data on teachers that is about to be made public is meaningless, which is why FEA joined in to enforce the public records exemption and prevent it from being published. The numbers to be released are subject to misinterpretation. They have not been put in their proper context.
- FEA fully supports teacher accountability. But assessments of teachers, like assessments of students, must be valid, transparent and multi-faceted. These value-added model calculations are none of these. We hope that The Florida Times-Union – and anyone else who publishes these numbers – makes it fully clear to its readers how little meaning these numbers have in determining the quality of an individual teacher.

$$y_{ti} = \mathbf{x}_i \boldsymbol{\beta} + \sum_{r=1}^L y_{t-r,i} \gamma_{t-r} + \sum_{q=1}^Q \mathbf{z}_{qi} \Theta_q + e_i$$

- The FEA fully supports teacher accountability, as no one wants an ineffective teacher in the classroom. But assessments of teachers, like assessments of students, must be valid, transparent and multi-faceted. These value-added model (VAM) calculations are none of these.
- Just look at that formula above. It is ludicrous to try to determine the value of a teacher using a formula that is comprehensible only to a small number of statisticians. With the problems that the DOE has been having with data on testing and school grades, we have little confidence in these complex figures used to determine a teacher’s evaluation.
- The numbers released by the DOE are subject to misinterpretation. They are mechanical calculations and have not been put in their proper context. The complex value-added statistical model is part of a highly complex accountability system. The Florida public is well aware of the ongoing problems with FCAT and school grades. The full accountability system must be examined and reimaged.
- Research has shown that even the most sophisticated and valid VAM measurements are limited in what they can measure. But Florida’s VAM formula is not valid; it is deeply flawed in practice.

- The two-year cumulative number includes data from 2009-10, before the law even went into effect, calculated after the fact with gaping holes of missing data and little or no roster verification.
- Nearly all teachers' VAM numbers are calculated according to students' FCAT scores, yet only approximately 35 percent of teachers teach students and subjects tested on the FCAT. So for 65 percent or more of teachers, the VAM does not even attempt to measure the teacher's actual teaching.
- The Legislature openly recognized this flaw in passing SB 1664, which requires future VAM scores to be based upon a teacher's actual students. But the two years of VAM data to be released by DOE does not take into account the new law, making all of the data meaningless.
- After two years, millions of dollars spent on the formula and countless hours spent in implementing SB 736, if these calculations are all the DOE has to show for its efforts, Floridians can now see what this effort has accomplished for Florida's students: Nothing. It's an expensive boondoggle that is taking money from the classroom and putting it in to the pockets of the private consultants contracted by DOE. Our students deserve better.
- Researchers have issued numerous warnings about basing teacher evaluations substantially on student test scores. The Legislature and the DOE have largely ignored these warnings.
- Most researchers agree that VAM is not appropriate as a primary measure for evaluating individual teachers. Reviews of research on value-added methodologies for estimating teacher "effects" based on student test scores have concluded that these measures are too unstable and too vulnerable to many sources of error to be used for teacher evaluation.
- Value-added models, taken by themselves, are not an adequate measure of overall educational quality. Like any other measure based on standardized tests, VAMs provide an incomplete view of students' knowledge, skills and dispositions. Standardized tests only assess a fraction of what teachers teach and students learn; VAM scores based on these "fractional" standardized test results should not be used as the singular assessment of a teacher's impact on student learning.
- Teachers' ratings are affected by differences in the students who are assigned to them. Statistical models cannot fully adjust for the fact that some teachers will have a disproportionate number of students who may be exceptionally difficult to teach students with poor attendance, who are homeless, who have severe problems at home, etc. Also, the model does not accommodate high performing students who "hit the ceiling" with near-perfect or perfect test scores and cannot show growth. For example, a teacher of gifted students may have a negative VAM score based on student learning growth even though all of his students consistently perform at the highest achievement level each and every year.
- Value-added models of teacher effectiveness do not produce stable ratings of teachers. Teachers look very different in their measured effectiveness when different statistical

methods are used. In addition, a given teacher may appear to have differential effectiveness from class to class, from year to year, and even from test to test. Researchers have noted that ratings are most unstable at the upper and lower ends of the scale, where many would like to use them to determine high or low levels of effectiveness.