

Basic Problems with the Ohio Teacher Evaluation System

Randy L. Hoover Ph. D.
(2014)

Preface

The Ohio Teacher Evaluation System (OTES) represents another waypoint in Ohio's march toward mesmerizing us into believing that school reform is going to create effective schools and effective educators to the benefit of all Ohioans. No matter that the entire accountability system has been carefully contrived to build political careers by funneling billions from public funds built by Ohio taxpayers into for-profit corporations through vouchers, charter schools, curriculum packages, standardized tests, and test materials. Everywhere, school reform is bully politics but nowhere more than in Ohio, as hardworking and dedicated teachers now stand defenseless in front of the firing squad of the Ohio Teacher Evaluation System.

The Ohio Teacher Evaluation System (OTES) will likely not be the last school initiative to come from Columbus, but it is clearly a major milestone in the long history of anti-public school and anti-teacher legislation and policy dating back to the Voinovich years. The time is long, long overdue for concerted and coordinated action to expose the hidden agendas and the fictions that have led to the indefensible absurdity of OTES. The mission to first and foremost educate the educators must become the prime directive for those who understand the subterfuge of school reform in Ohio.

Educating the educators is about empowering the now nearly-powerless public school teachers to recapture the leadership of their unions to then regain the understanding and support of Ohio's citizens. *The teachers of Ohio want accountability* but in the form of authentic accountability not in the form of pseudo accountability as is exemplified by the Ohio Teacher Evaluation System. To these ends, there is no better place to expose the insidious nature of Ohio's school reform than in the expose' and critique of the Ohio Teacher Evaluation System.

The two basic elements consisting of 1) *The Teacher Performance section* and 2) *The Student Growth Measures section* are the dual measures for the evaluations. To fully deconstruct and expose the flaws within OTES is impossible within the limits of what time and resources permit. Because of the importance of understanding OTES and because of the complexity of its metrics, I strongly recommend reading the related articles found on *The Teacher Advocate* homepage. This discussion of OTES will be revised as issues arise and questions are submitted. Identifying and summarizing some of the key issues of OTES may allow readers themselves to further explore the full extent of the failure of OTES to be a well-reasoned and credible form of teacher evaluation.

The Teacher Performance Section

The overarching issue with the Teacher Performance section of OATES is the professional messiness of the entire component. I use the notion of professional messiness because there is so much left to the subjective interpretation of everyone involved that precision of meaning and any sense of internal consistency are lost. Yet dedicated teachers are asked to surrender to the convoluted procedures with their own specific narratives for conferencing, observations, self-improvement plans, and classroom performances.

The linking of the teacher evaluation rubric with the Ohio Standards for the Teaching Profession is such that given the high-stakes nature of OTES, the rubric becomes the target for the performance *more in terms of how the teacher sees the subjectivity of the evaluator's interpretation* of the rubric as being the goal for performance rather than the standards themselves or even what they know is best for their students.

This is professionally messy because both the standards and the rubric lack specific meanings. Items have what is referred to as "high inference levels." *Inference level* is a concept taken from the language of scientific research where it is used to address the precision and the certainty of an observation or variable. The higher the inference level, the more there is left to the imagination. The professional

messiness is found in the fact that the higher the inference level of an OTES rubric item, the more the performance is open to the uncertainty of the teacher (the performer) and, more importantly, to the interpretation of the evaluators (the judges).

The teacher evaluation rubric requires a highly subjective judgment because the behaviors to be observed are, for the most part, extremely high inference—they leave much to the uncertainty and imagination. Because of this, it is the type of evaluation mechanism that gets those being evaluated geared to try to please the evaluator rather than to embrace and engage actual elements of effective teaching. Survival trumps all else.

For the classroom teacher at any grade level, the natural fear is looking at the performance measure of OTES as a way to arbitrarily lower the teacher evaluation. If a teacher is rated lower than expected in this section of OTES, the teacher is at the mercy of the principal or other evaluator *for their final rating because there is no appeals process embedded in OTES. The absence of an objective appeals process completely destroys the integrity of OTES.* The only grounds for appeal would be limited to issues of violations of specified OTES procedures or timeline errors. Not only does not having any appeals process create a heavy-handed sense about OTES and destroy its integrity, it also violates well-established standards for educational accountability.

Many educators do not know that there are two well-established sets of standards developed by recognized experts in the psychometrics of testing and accountability systems. *The Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) represents the standards for tests and testing developed by a consortium of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). This set of standards deals primarily with tests and testing themselves.

To address critical issues of educational accountability systems and the assessments and evaluations used in those systems, *The Standards for Educational Accountability Systems* were produced for the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) by a consortium comprised of the same professional groups listed above. The CRESST standards consist of 22 criteria that represent rigorous guidelines for all educational accountability systems. The test standards are widely recognized as the most authoritative statement of professional consensus regarding expectations on matters of validity, fairness, and other technical characteristics of tests (Baker & Linn, 2002). Adherence to all of the 22 standards is a requirement for any given accountability system to be credible. Regarding appeal procedures, CRESST gives us:

16. *Appeal procedures should be available to contest rewards and sanctions.*

Comment: Extenuating circumstances may call the validity of results into question. For example, a disturbance during test administration may invalidate the test results. *Also, individuals may have information that leads to conflicting conclusions about performance. Appeal procedures allow for such additional information to be brought to bear on a decision and thereby enhance its validity.* (p. 4) (Italics added.)

Conceptually, OTES has a major problem in terms of the validity of the evaluation. The fact that there is no universal, agreed-upon, specific definition of effective teaching adds to the messiness of OTES. While the state offers claims about effective teaching in the document “Ohio Teacher Evaluation System Model” (Ohio Department of Education, 2012), *nowhere is there any discussion or specification of the outcome that teacher effectiveness serves.* There is no discussion of what the purpose of going to school actually is. Therefore, effective teaching has no benchmark for evaluating (judging) its outcome because the state does not provide any meaningful definition of the outcome. It is logically impossible to judge effectiveness of any kind if the outcome, the ends, are completely absent. It is tantamount to telling teachers that we won’t tell you what effective teaching is, but you damned well better do it!

Indeed, this kind of meaningless, boiler-plate language is found throughout the Ohio Department of Education (ODE) discussion of teacher effectiveness. On page 3 of the OTES Model, for example, it is written:

After conducting extensive research, the following definition of teacher effectiveness was developed by educational practitioners in Ohio and is reinforced by Ohio's Standards for the Teaching Profession. Clearly the research supports the direct connection between effective teaching and high student achievement.

At first glance, nothing seems amiss until we realize that without specifically defining "student achievement," the statement is entirely meaningless. Clearly, the notion of student achievement implies a teaching outcome, the ends for the means of effective teaching and schooling. There are a number of ways student achievement has been defined and discussed throughout our history, most of them at great variance with each other. For each different perspective on achievement outcome, teacher effectiveness would also likely be different.

Another example of the frustrating messiness created in the OTES Model discussion of teacher effectiveness is seen in one of the seven statements on page 3 where it is stated that effective teachers, "Plan and deliver effective instruction that advances the learning of each individual student." The statement is obviously circular—effective teachers teach effectively. The statement tells the reader absolutely nothing because it is purely political rhetoric to make the state's reform efforts sound good.

The next claim in the same section says effective teachers "Create a learning environment that promotes high levels of student learning and achievement for all students." This begs the question of what is meant by learning and achievement. Their statement also implies learning and achievement are two different concepts. Clearly, this is a distinction without difference, further indicating the lack of thought put into the document.

The entire state discussion of effective teaching and achievement is a ruse used to distract any reader from the reality that scores on standardized achievement tests are the sole purpose of teaching and schooling in Ohio. The teacher performance side of OTES is a meandering sidebar of arbitrary requirements that are extremely time consuming and worrisome for our classroom teachers. This side of OTES is devoid of legitimate scholarly research and completely absent of any well-grounded theory of teaching.

But given the nature of Ohio's entire education accountability system's complete reliance on standardized test scores, the *implicit* purpose of schooling in Ohio is to get high student test scores. Logic then dictates that the effective teacher is defined as the teacher whose students get high scores on the tests. Therein, we are firmly stuck in the messiness of OTES and the most narrow and limited *de facto* definition of the purpose of schooling possible—high scores on standardized tests.

The Student Growth Measures Section

The Student Growth Measures section of OATES is perhaps the best example of metrics-gone-wrong that can be found within any education accountability system. With its trinity of achievement test options, the student growth section accumulates a number of significant testing violations. In doing so, this section of OATES locks teachers (and subsequently administrators) into a full-blown model of pseudo accountability¹ that holds them responsible for very powerful forces and factors that are completely beyond the realm of their professional control.

Even before discussing the critical flaws of validity, limited teacher effects on students, misuse of value added, and the complete impossibility of using locally determined measures to determine teacher effectiveness, the Student Growth section overtly violates another very important accountability standard. Returning to the issue of Ohio's blatant violation of a key standard for holding educators accountable for student performance, we turn to Standard 15 of the CRESST standards:

¹ See "Pseudo Accountability vs. Authentic Accountability" to understand this aspect of OTES.

15. *Stakes for accountability systems should apply to adults and students and should be coordinated to support system goals.*

Comment: Asymmetry in stakes may have undesirable consequences, both perceived and real. For example, *if teachers and administrators are held accountable for student achievement but students are not, then there are likely to be concerns about the degree to which students put forth their best effort in taking the tests.* Conversely, it may be unfair to hold students accountable for performance on a test without having some assurance that teachers and other adults are being held accountable for providing students with adequate opportunity to learn the material that is tested. Incentives and sanctions that push in opposite directions for adults and for students can be counterproductive. They need to be consistent with each other and with the goals of the system. (p. 4) (Italics added.)

It is not hard to imagine how the effects of this asymmetry can vary widely across grade levels and subject areas and negatively affect many classroom teachers. Ohio's reliance on standardized testing in general and testing's role in OTES provides no consideration of the very real effects asymmetry will have on teacher evaluations. Just as with the accountability system as a whole, the Ohio Teacher Evaluation System makes the assumption that all students have identical attitudes toward their school, the standardized tests, and their teachers. As is discussed in the [paper on value added](#) on this site, the peril of student apathy or even antagonism teachers face from students taking tests that significantly affect the teacher's career is, in many cases, daunting.

Another even greater flaw in OTES is the failure to incorporate the research that examines teacher effects on student achievement. With OTES, just as with the entire accountability system, the overwhelming research evidence that shows teachers have only slight impact on student achievement is ignored. Likewise, the extreme difficulty of determining that impact with any precision is also ignored. The state operates as if the teacher is the single most powerful force in determining student academic performance when we know for a fact that is not the case, especially when it comes to test performance. In a direct response to the fallacious assumptions accountability systems make about teacher effects, Scott McLeod (2013) writes:

Another issue worth noting is that even if teacher effects could be teased out, decades of peer-reviewed research show that teachers only account for about 10% of overall student achievement (give or take a few percentage points). Another 10% or so is attributable to other school factors such as leadership, resources, and peer influences. The remaining 80% of overall student achievement is attributable to non-school factors such as individual, family, and neighborhood characteristics. A few exceptional 'beating the odds' schools aside, these ratios have remained fairly stable (i.e., within a few percentage points) since they were first noted by the famous Coleman Report of the 1960s. Given the overwhelming percentage of student learning outcomes that is attributable to non-teacher factors, it is neither ethical nor legally defensible to base teacher evaluations on factors outside of their control. (McLeod, 2013)

McLeod's commentary supports my own research findings (Hoover, 2000, 2008) as to the limited extent to which teachers can have academic impact against the lived experience of the student as measured by achievement tests. However, the primary point is that good science clearly indicates *it is simply impossible to accurately determine the impact teachers have on their students because the issues of test validity, high measurement error, and the forces and factors outside the control of the teacher are so powerful in and of themselves irrespective of the quality of the teacher.*

The last item above, the forces and factors outside the control of the teacher, is the gateway into the realm of high-stakes pseudo accountability—the place of no return for those teachers in schools with lower socio-economic student populations.

Value Added as a Gain Score Indicator of Teacher Effectiveness

OTES is completely test driven in the Student Growth Measures section. Aside from the very serious flaws itemized so far, the problem for teachers getting any kind of fair and valid evaluation is rendered impossible with the addition of student gain scores, primarily in the form of value-added scores, being used to categorize teacher performance on the Ohio 1- to 5-point teacher effectiveness scale. The student gain score component is required in all three categories with a value-added component of OTES being apparently derived from the first two of the following three sources: 1) *The mandated standardized achievement test scores (if tests are given at that grade level and subject area,* 2) *Vendor-produced standardized tests for grade levels and subject where such tests are available,* and 3) *Teacher-made tests for grade levels and subject areas where no standardized vendor tests exist.*

While I strongly advise the reader to see the section on [value added](#) on this site for greater detail and explanation of value-added issues, there are some important observations about value added that must be also made here in dealing with OTES. The degree of suspicion about value added that I have encountered talking with teachers is monumental, to say the least. For the dedicated classroom teacher typical of Ohio's public schools, the lack of understanding the concept of value added is replete just as is the threat implied by value added. Having studied value added for many years, I can assure our classroom teachers that there are very sound reasons to feel threatened by the inclusion of value added in any form of teacher evaluation system.

Ohio's value-added model (VAM) has several fatal flaws that destroy any credibility of the results for either the individual educator or for the building and district levels. The most damning flaw is that value added uses data from standardized tests that lack test validity. Whether the value-added score comes from the typical Ohio Achievement Test or any other vendor-created achievement test, *as long as those test scores show statistically high correlations with the socio-economic quality of the lived experience of the students, they cannot claim to have test validity.*

The important point here is that if the primary data (the test scores) fed into the value-added metrics are not demonstrably valid (and absent all socio-economic effects), then any value-added results are equally not valid. The old adage of "garbage in, garbage out" applies fully to the value-added metrics. The reality for classroom teachers is that *Ohio's use of value added fails to provide any indication whatsoever as to the actual effectiveness of any teacher.*

However, it is important to understand that value added in and of itself is neither bad nor good in terms of its results. Making a value judgment about VAM depends on what it is used for and how it is used. There are uses for VAM in manufacturing and agriculture that are highly valid, and the results are, therefore, useful and credible.

At the root level, where school reform targets students by requiring them to pass achievement tests, the notion of measurement error takes on even greater relevance in the form of each score being only as accurate as the confidence interval of the score as a function of test reliability. Baker and Linn (2002) comment on how this may play out and mislead:

The use of performance standards to determine whether a student is proficient or not (passes or fails) reduces test scores to a dichotomy. Measurement error that is associated with any test score results in classification errors. That is, a student whose true level of achievement should lead to a passing score earns a score that is below the passing standard and vice versa. Valid inferences about student proficiency are undermined by measurement errors that result in misclassification of students. Hence, it is critical that the probability of misclassification is evaluated and the information is provided to users of the performance standards results. The precision of test scores can be enhanced by increasing test length. As Rogosa [1999] has shown, however, even tests that have reliability coefficients normally considered to be quite high (e.g., .90) result in substantial probabilities of misclassification. For example, if the passing standard is set at the 50th percentile for a test with a reliability of .90, the probability is .22 that a student whose true percentile rank is 60, and who therefore should pass, would score below the cut score and

therefore fail on a given administration of the test. Even a student whose true percentile rank is 70, a full 20 points above the cut score, would have a probability of failing of .06 [Rogosa, 1999]. (Baker and Linn, 2002, p. 9)

This example from Baker and Linn shows what happens at the root level of the accountability systems in terms of confidence intervals. The confidence interval for a test score is the calculated range of the score, not a precise, single value as is assumed by the Ohio accountability system. But whether the tests in the example are no-stakes, low-stakes, or high-stakes makes a tremendous difference. Typically, test scores like those in the Baker & Linn (2002) example *are the primary sources of data for high-stakes categorization of teachers,² individual schools, and districts.³*

When student test scores are used as the dominant source of data for overall accountability systems, the likelihood of any accuracy and, therefore, clear meaning is lost. Value-added models are the exemplars of this. Mathew Di Carlo, senior research fellow at the Albert Shanker Institute, writing about the work of Sean Corcoran (2010) of the Annenberg Institute for School Reform, notes how using value-added models greatly amplifies the role of measurement error:

Interpreting a teacher's VAM score without examining the error margin is, in many respects, meaningless. For instance, a [recent analysis](#) of VAM scores in New York City shows that the *average* error margin is plus or minus 30 percentile points. That puts the "true score" (which we can't know) of a 50th percentile teacher at somewhere between the 20th and 80th percentile—an incredible 60 point spread (though, to be fair, the "true score" is much more likely to be 50th percentile than 20th or 80th, and many individual teacher's error margins are less wide than the average). If evaluation systems don't pay any attention to the margin of error, the estimate is little more than a good guess (and often not a very good one at that). Now, here's the problem: Many, if not most teacher evaluation systems that include VAM—current, enacted or under consideration—*completely ignore this*. (Di Carlo, 2010)

Out of OTES being an example of metrics-gone-wrong, by far, the most egregious aspect of it all is the requirements for teacher-made tests to be used for grade levels and subject areas where no standardized vendor tests exist. Assessments of any kind must adhere to the well-established principles of test construction as well as to the procedures and metrics that evidence validity and reliability. In the case of the teachers who teach in areas where no vendor tests yet exist, the requirement to produce Student Learning Outcomes (SLO) and to produce appropriate pre assessments and post assessments that will determine comparative achievement gains is overwhelmingly complex. The process is extremely labor intensive and incredibly time consuming. The process also requires advanced expertise in psychometrics, something very, very few educators have had the opportunity to learn. That ODE expects the pre assessment to be different from the post assessment more than doubles the work. The teacher now needs to do complete item analyses and do the calculations for the determinations of evidence for the validity and the reliability for two tests, all of which require extensive field testing.

It would take many hours to list and explain the problems that exist for anyone to create this Student Learning Outcome assessment correctly—with due attention to the mathematical rules and procedures necessary to have appropriate SLOs with fully valid and reliable pre assessments and post assessments. There is not enough time to list and explain the issues involved in doing the job correctly by adhering to the established standards of the fields of psychometrics. (I would welcome anyone with a background in this area to submit a written overview of what is specifically involved that we could publish on this website.)

² A recent analysis of VAM scores in New York City shows that the average error margin is plus or minus 30 percentage points. That puts the "true score" (which we can't know) of a 50th percentile teacher at somewhere between the 20th and 80th percentile—an incredible 60-point spread (Di Carlo, 2010).

³ In the 2000 study, when I controlled for the lived experience factor (socioeconomic status), I found many apparently high-performing districts to be performing below their state ratings and many low-performing districts actually performing far above.

Suffice it to say that there are very few people with the technical knowledge needed to produce these assessments correctly. To expect any classroom teacher to do this is absurd even with the information found on the ODE website. The degree of difficulty originating from the labyrinth of explanations and directions on the ODE website pages alone is extremely confusing and equally frustrating. More importantly, the knowledge necessary for the teacher to do the job is largely absent from ODE's website. Putting myself in the shoes of a teacher, I went into the ODE site pages dealing with assessments in an attempt to find out how to create appropriate pre and post assessments for SLOs. After a considerable amount of effort, I was finally able to locate a PDF "Appendix C. Guidance on Selecting Assessments for SLOs⁴." But I found this document would be of little or no help to any teacher needing to understand and apply the psychometric knowledge necessary for developing valid and reliable tests. Sadly, the ODE document actually trips over itself when it confuses reliability with validity (p. 7) and then completely misuses the concept of item analysis (p. 4).

Most conspicuous by its absence in the ODE guidance document is any mention of the role of field testing in verifying formal item analysis, establishing the validity, and evidencing the reliability of the assessments. Field testing is an absolutely unavoidable requirement for this type of test development. For example, it is impossible to do an item analysis such as determining item level of difficulty and discriminating power without concerted field trials. Likewise, determination of test reliability, especially determination of the measurement error, is required if the test scores are to have any utility and credibility at all. In the case of pre and post assessment instruments, only field testing can assure that both tests are equivalent in assessing what they are supposed to assess. For any classroom teacher to employ SLO assessments without the appropriate rigorous field testing places the teacher in a very high-stakes situation with no possibility of getting any valid or credible data on student growth whatsoever--none. Thus, the evaluation of the teacher on the student gain score side of OTES is left to random-chance results. The failure of ODE to lead the way in educating teachers about this unavoidable procedural requirement in its own "guidance" document is grossly negligent and completely unethical.

The reality is that this primary ODE guidance document is of no use to a teacher in creating truly reliable and valid assessments. *For teachers who must go the SLO route for their evaluations, not having expertise or access to expertise puts them in peril of getting poor evaluations due to the nature of their tests, not the nature of their teaching effectiveness* because they are not told *what* the specific requirements are for good tests nor are they told *how* to create them. What our teachers are given on the ODE site is incomplete, inaccurate, and imprecise.

When I was first told that teachers in classes with no vendor tests were being required to design the SLO-Pretests-Posttests themselves, my response was that it simply can't be done. My meaning was that it can't be done properly by classroom teachers, through no fault of their own. It takes a great deal of study to gain the knowledge to do it properly. Determining the discriminating power of test items and doing item analysis alone require advanced mathematical and conceptual understandings. How many classroom teachers are well versed in calculating standard error of measurement or determining the confidence interval for the scores on the tests they will create? Indeed, I know of only a very few even in higher education who are knowledgeable enough to perform these kinds of highly-complicated assessment activities.

The requirements of OTES in this SLO dimension put classroom teachers squarely in harm's way. Teachers are placed in situations where the results of their work, if not done precisely and with complete understanding of relevant concepts of psychometrics, can well put them in a low or failing category for their evaluation results. The gain scores for these teachers will be vastly more indicative of their lack of knowledge about test development than their effectiveness as teachers. How are evaluators to know if the teacher's results are not simply the result of not knowing the ins and outs of good test development?

⁴ <http://education.ohio.gov/getattachment/Topics/Teaching/Educator-Evaluation-System/Ohio-s-Teacher-Evaluation-System/Student-Growth-Measures/Student-Learning-Objective-Examples/guidance-selecting-assessments.pdf.aspx>

Again, the conspicuous absence of an appeals process speaks volumes about the negative attitude toward teachers held by the Governor, the State Legislature, and ODE. Similarly, the failure of our unions to fully and candidly inform their memberships of this peril speaks volumes to their legacy of indifference in terms of being teacher advocates. It is one thing for the unions to give opponent testimony for the SLO part of the OTES legislation and quite another to not fully explain the critical issues and risks to their memberships, not to mention the legislature and citizens of Ohio.

The multiple fatal flaws of OTES notwithstanding, we need to return to the big picture and the singularly most important false assumption that underlies the entirety of OTES. Just as previously seen in the McLeod (2013) summary of the research on teacher effects, Rothstein (2010), writing for the highly respected Economic Policy Institute (EPI), speaks directly to the most fundamental flaw in teacher evaluation systems such as OTES:

It has become conventional in educational policy discussion to assert that “research shows” that “teachers are the most important influence on student achievement.” There is, in fact, no serious research that shows any such thing. The assertion results from a careless glide from “teachers being the most important in-school influence,” to teachers being the most important influence overall. But because school effects on average levels of achievement are smaller than the effects of families and communities, even if teachers were the largest school effect, they would not be a very big portion of the overall effect. A child with an average teacher who comes from a literate, economically secure, and stable family environment will, on average, have better achievement than a child with a superior teacher but with none of these contextual advantages.

Further, this factual reality of Ohio’s test validity failure subsequently and directly destroys the integrity of any and all its uses as data inputs for additional analysis such as Ohio’s building or school ratings and all value-added performance claims. The inescapable conclusion is, therefore, that Ohio’s teacher and administrator evaluation systems are in no way, manner, shape of form indicative of professional effectiveness. We need to understand that any educator evaluation element based upon the test scores is a sham beyond any doubt. It is *pseudo accountability* in extreme.

Conclusion

Today, the only route to finding a fair, valid, reliable, and reasoned educational accountability system for Ohio is to begin by clearly understanding the monumental failure of the current system. Whether we are addressing the system overall or only the component of the Ohio Teacher Evaluation System, we need to know and be able to articulate the difference between pseudo accountability and authentic accountability systems. The prime accountability directive for all of us is twofold:

- 1) *As professional educators, we must actively seek being held accountable for those aspects of the profession within our professional control.*
- 2) *We must emphatically reject any and all elements of any system that holds educators or our public schools accountable for anything not clearly within the scope of professional control.*

The prime accountability directive must be the benchmark for judging the worthiness of any law or public policy relating to our accountability. The prime accountability directive must also be the benchmark for all stakeholders to apply in judging the credibility of Ohio’s educational accountability system. Just as for educators, if stakeholders want education accountability, they must insist upon a system that is demonstrably credible.

From the standpoint of OTES, the prime accountability directive is clear. The entirety of OTES is holding public schools, their teachers, and their principals accountable for forces and factors that are completely beyond their professional control. Therefore, OTES must be rejected as wholly unacceptable no matter the political consequences. Our professional association leadership must be held responsible for resolutely working to end the morally wrong and logically indefensible evaluation system imposed on Ohio’s teachers.

References

- AERA, APS, NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baker, E., & Linn, R. (2002). *Validity issues for accountability systems*. Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Center for Research on Evaluation, Standards, and Student Testing (2002). *Standards for educational accountability systems*. Los Angeles: UCLA.
- Corcoran, S. (2010). Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. Retrieved from <http://www.scribd.com/doc/37648467/The-Use-of-Value-Added-Measures-of-Teacher-Effectiveness-in-Policy-and-Practice#download>
- Di Carlo, M. (2010). The war on error. Retrieved from <http://shankerblog.org/?p=1383>
- Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). Developing and selecting assessments of student growth for use in teacher evaluation systems (extended version). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from <http://www.cse.ucla.edu/products/policy/TestScoresTeacherEval.pdf>
- Hoover, R. (2000). *Forces and factors affecting Ohio proficiency test performance: A study of 593 Ohio school districts*. Retrieved from <http://people.yosu.edu/~rlhoover/OAT-OGT/index.html>
- Hoover, R. (2004). *(Re)understanding educator accountability: Pseudo vs. authentic accountability*. Retrieved from <http://people.yosu.edu/~rlhoover/OAT-OGT/index.html>
- Hoover, R. (2008). *A Re-examination of forces and factors affecting Ohio school district OAT and OGT performance*. Retrieved from <http://people.yosu.edu/~rlhoover/OAT-OGT/index.html>
- McLeod, S. (2013). Value added measures: VAM. Retrieved from <http://dangerouslyirrelevant.org/resources/value-added-measures>.
- Rothstein, R. (2010). *How to fix our schools: It's more complicated, and more work, than the Klein-Rhee 'manifesto' wants you to believe*. Economic Policy Institute, Issue Brief #286. Washington, DC: EPI