

# Testimony for House Bill 5223

by Doug Greer, Ottawa Area ISD, a member of the West MI Talent Triangle

In order for our students to reach their fullest potential, we need to have an educator evaluation system that provides timely feedback on instruction and the appropriate support structure to help educators have impactful professional development. This legislation contains many positive components that help us accomplish this goal. We appreciate the bill sponsors maintaining four categories (highly effective, effective, minimally effective, ineffective) that identify those teachers who truly exceed all expectations. We also appreciate maintaining a 25% growth requirement for three years as we transition to a new assessment system. **Suggestion: Allow districts to determine the 25% until the percentage is increased in 2017/18; then prescribe a minimum portion for state level assessments.**

## What is the purpose for state level assessments?

“A critical task for policy makers is to answer explicitly the first question posed above—‘What do I want to learn from this assessment’ — and then find or develop a set of assessments that best fits that purpose. We see three general classes of purposes for assessments: *instructional, evaluative, and predictive*. Within each general class, there are myriad specific purposes.” (Perie, et al, p. 4)  
“Given constrained resources, it is no wonder that educational leaders are tempted to use a single assessment system for as many purposes as possible. Unfortunately, one of the truisms in educational measurement is that when an assessment system is designed to fulfill too many purposes—especially disparate purposes—it rarely fulfills any purpose well.” (Perie, et al, p.5)

Therefore, it is critical to define the primary purpose of a quality state standardized assessment system to validly and reliably **evaluate student achievement and growth** with respect to Michigan’s state-adopted standards. Assessments can only have one primary purpose, if we were to prioritize those purposes in an attempt to accomplish more than one purpose, we must articulate evaluation as the highest priority. We would then expect the assessment to inform instruction (2nd) and we would hope that we could also predict college readiness (3rd).

## What is the importance of assessment in educator evaluations (HB 5223)?

### (G) PROVIDE AN ITEM LEVEL RESPONSE SUMMARY (pg. 15)

The type of reports produced rely heavily on the design and primary purpose of the assessment: evaluative, predictive or instructional. In addition, the delivery method chosen will also vary the type of reports. Item level response summary reports are limited to non-adaptive assessments which limit the ability of the assessment to accurately measure the highest and lowest achieving students. Also, test security plays a role in whether actual items are released or sample items. In order to inform instruction, it is critical that the reports align to the state-adopted standards or small groups of standards, known as clusters in the Common Core State Standards for ELA and mathematics. **Sample language: Provide a standards based response summary of state-adopted standards with sample items for each standard or standard group.**

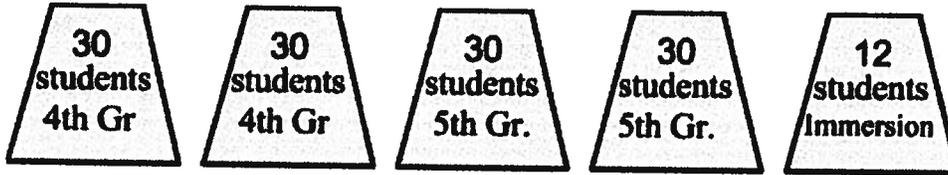
### (ii) PROVIDE STUDENT GROWTH (AND ACHIEVEMENT) DATA FOR EDUCATORS ON THE STATE-MANDATED ASSESSMENTS. (pg. 16)

How we measure student growth and achievement is truly one of the most important topics not limited to this bill alone. Holding teachers accountable for a group of students is parallel to holding buildings accountable to usually a larger group of students:

High School Math Teacher may have 5 classes with 132 students who s/he is held accountable for teaching the state adopted high school math standards.



Likewise, a small elementary may have 5 classes with 132 students that the building principal and the building are held accountable based on state testing.



Therefore, we want to be extremely cautious about not repeating some of the same errors found in our current accountability system for both small and large buildings.

One of the major concerns with our current system is the ability for a mathematician to accurately predict the ranking or simply the achievement solely based on poverty. We should NOT be able to accurately predict an evaluation simply based on demographics such as percent of students who qualify for free or reduced lunch.

The Mackinac Center (Spalding, 2013) highlighted the predictability of poverty as a major concern for Michigan's current accountability system (Top to Bottom Rankings) and compared them to other states such as Ohio, Indiana, and Arizona. The graph to the right illustrates that Arizona who considers how buildings perform when compared to similar cohorts has the lowest predictability with roughly 12%. Ohio utilizes a different model that Wayne Kuipers has promoted through SAS-EVAAS, known as a Value-Added Model (VAM) resulting in a predictability of 35%. Michigan who does not currently use peer grouping, VAM, or other methods of measuring achievement and growth has an unacceptable predictability of 55-60%.

*Will an emphasis on growth fix this?*

**Why does it matter that we make a consideration for similar groups of aggregated students when holding buildings or educators accountable?**

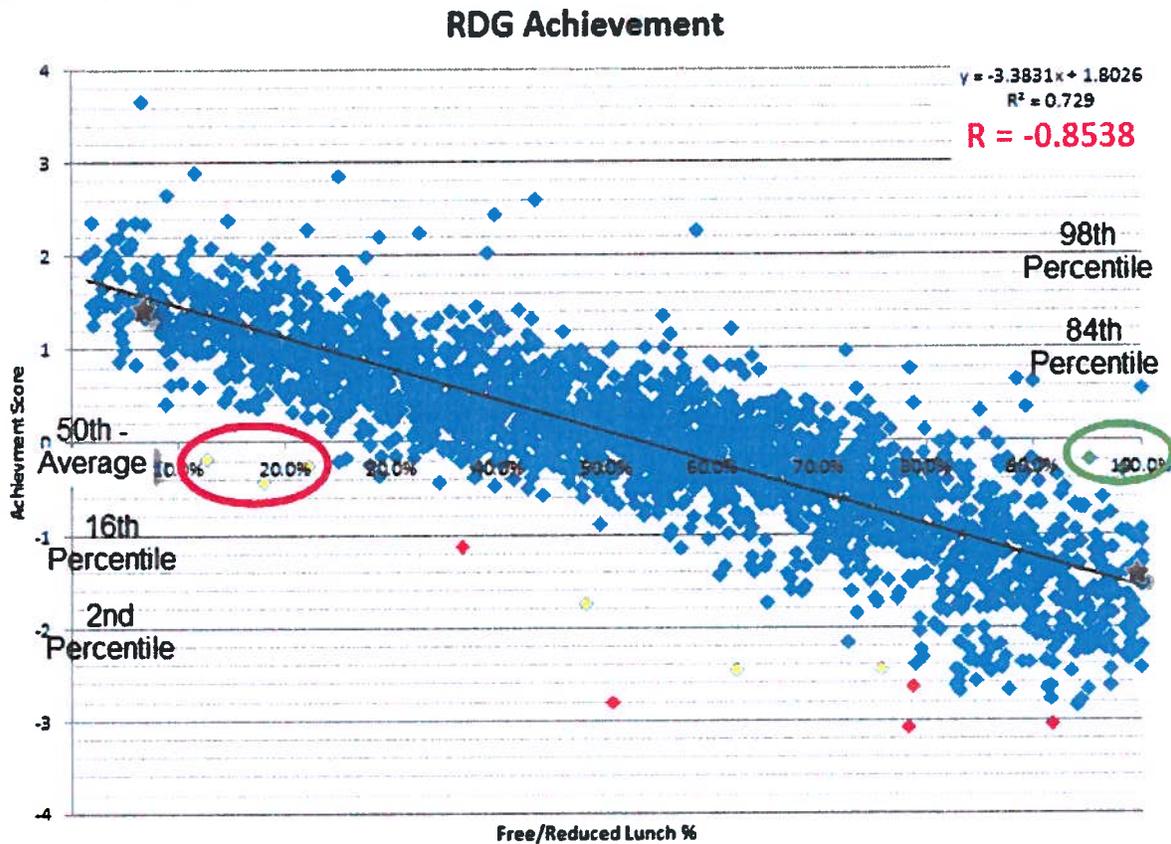
**Graphic 4: School Accountability Systems and School Poverty Rates**

State	School Year	Percent of Ranking Explained by Poverty ( $R^2$ )
Michigan	2012-13	55%
	2011-12	56%
	2010-11	61%
Florida	2012-13	47%
	2011-12	47%
Wisconsin	2011-12	42%
Oklahoma	2011-12	35%
	2010-11	35%
Ohio	2011-12	35%
	2010-11	35%
Maine	2011-12	32%
Indiana	2011-12	30%
Arizona	2012-13	12%
	2011-12	13%
	2010-11	11%

Source: See "Measuring A Data Source for School Accountability Systems and School Poverty Rates"

Imagine two teachers with very different circumstances but very similar student scores. Both teachers have student scores that when averaged fall at the state average. Therefore, the assumption is that these students are very similar and both teachers are simply average. However, Mr. Rich receives a fairly poor observational score, though the state assessment brings his ranking back up because his students score at the state average. On the other hand, Mrs. Jones has a very high observational score, unfortunately her students are only performing at the state average so she is unable to achieve "Highly Effective."

Now look at the graph below, the three yellow dots circled in red represent Mr. Rich (real data from 2 charter schools and one traditional public). The two green dots circled in green represent Mrs. Jones (real data from one charter and one traditional public). When we look at a peer comparison, we can clearly see that Mr. Rich is performing significantly below his peers and Mrs. Jones is significantly above her peers. How do we assure that Mr. Rich is not protected by a new evaluation system? **Achievement and Growth data from state assessments shall not have a strong correlation to poverty.** (Sample range may based on other states as seen earlier with a possible predictability ( $R^2$ ) less than 35%)



**References:**

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). Role of Interim assessments in a Comprehensive Assessment System. Aspen Institute pp. 4-5.

Spalding, A. (2013). Michigan's Top-to-Bottom Rankings: A Measure of School Quality or Student Poverty. Mackinac Center for Public Policy p. 7

