

Why New York State's Growth Model Shouldn't be used to Evaluate Teachers



FEBRUARY 2015

INTRODUCTION —

Recently, there has been a great deal of discussion about the validity of the Annual Professional Performance Review Process in New York state. The results for 2013-14 showed 1 in 100 New York state teachers have been classified as ineffective through the Annual Professional Performance Review (APPR) process. The Governor and former Education Commissioner King indicated that the number of “ineffective” teachers is too low. No one has been able to point to any rational reason for why the number should be higher. There is no research that indicates that some specific percent of educators are incompetent. In New York, with our high standards for entering the profession, we have every reason to expect the numbers of ineffective teachers to be low.

The only rationale used by the Governor and former Commissioner King is the performance of students on the state common core assessments. This rationale is used despite the fact the Legislature and Governor have taken action to ensure these tests are not used against students and the State Education Department (SED) has admitted they did not implement the changes properly.

Everyone wants competent teachers in every classroom. To ensure this happens, the process of developing effective teachers begins long before the first evaluation and continues throughout their career. New York state requires teachers to earn initial certification by receiving a bachelor's degree, spend a semester of supervised field work (student teaching), pass no fewer than four certification exams, including the completion of the required performance-based assessment (edTPA).

In order to maintain certification and receive professional certification, teachers have at least 3 years of probationary teaching experience, including one year of mentoring and must earn a Master's Degree within 5 years.

Once they obtain their professional certificate, they must accrue 175 hours of additional professional development every five years to maintain their certification and must be evaluated annually.

In the first five years of teaching one in four exits the profession in New York state. Those educators who meet all the requirements are deemed highly skilled by the New York State Education Department and granted professional certification.

Some have suggested that the way to address this uninformed perception of “under classification of incompetent educators” is to put greater emphasis on the State Provided Growth Model. The growth model produces scores for educators based on how 4-8 ELA and math teacher’s students perform on state assessments in comparison to their peers. The growth model has proven to be harmful to the teaching profession, creating competition when collaboration is the desired approach, and this misguided recommendation will result in greater damage without any benefit to improve student learning.

In this document we will use the Standards for Educational and Psychological Testing, the same standards used by NYSED in the development of their state assessments, to bring to light the fact that the NYS growth model fails to meet the standards in multiple areas and therefore should not be used for teacher evaluation. The Joint Committee on the Standards for Educational and Psychological Testing is the authority on criteria and standards for the development and evaluation of tests and testing practices. The intent of these standards is to “promote sound testing practices and to provide a basis for evaluating the quality of those practices.”

Fact: Based on the SED description of the growth model, if EVERY student in grades 4-8 was proficient on the New York state ELA and Math Assessments 6% of the state’s 4-8 grade teachers, of ELA and math, would be classified as ineffective in the state provided growth model.

Standard: “Those who mandate the use of tests in policy, evaluation, and accountability contexts and those who use tests in such contexts should monitor their impact and should identify and minimize negative consequences.”*

The state provided growth model does not meet the standard

Using student test scores as part of the evaluation system is required by the NCLB waiver and is touted as an “objective” measure but for those teachers that receive state provided growth scores in New York state (grade 4-8 ELA and math teachers) it is also a relative measure. In a growth model, a teacher’s rating depends on how well they do compared to other teachers. By design, it is not a system where every teacher can be effective. There will always be winners and losers. For all other teachers (approximately 80%) the mandated student learning objective (SLO) process is on a fixed scale which does not limit the number of teachers that can reach any of the effectiveness categories. This unfairly treats 4-8 ELA and math teachers differently than all other teachers because they are not **all** able to reach an effective score. Teachers of special populations and economically challenged schools can be adversely affected by the model, dis-incentivizing teachers to work with these high-need populations.

One option that has been made available for districts to use for educators in subjects with no state assessment is the use of a school-wide group or team results based on state assessments – what this means is that teachers of subjects other than those being tested by state assessments would be measured based on student outcomes on assessments in classes the teacher did not teach – that measurement would then be used as part of the individual teacher performance review.

Fact: Neither of the state’s technical reports, for the assessments or the growth model, contains any indication that the developers of the state’s assessments and/or growth model have evaluated whether these assessments are in fact suitable for teacher evaluation.

Standard: “Evidence of validity, reliability and fairness for each purpose for which a test is used in a program evaluation, policy study or accountability system should be collected and made available.” Tests should not be used for multiple purposes unless there is evidence the test is valid for other use, meaning the specific inferences made from test scores is appropriate, meaningful, and useful.*

The state provided growth model does not meet the standard

Assessments are designed for specific purposes. According to the **New York State Testing Program 2013: English Language Arts Mathematics Grades 3–8 Technical Report**, “the Grades 3–8 Common Core ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State Common Core Learning Standards in ELA and Mathematics, respectively.” Neither of the technical reports, for the assessments or for the state provided growth model, provide evidence that supports the validity, reliability or fairness of using these assessments for educator evaluation.

Fact: According to SED data, the state provided growth model identifies a higher percentage of teachers in high needs schools as ineffective or developing than in low needs schools but the other APPR measures do not show a dramatic difference in the way educators perform by need category.

Standard: “Reports of group differences in test performance should be accompanied by relevant contextual information, where possible, to enable meaningful interpretation of the differences. If appropriate contextual information is not available, users should be cautioned against misinterpretation.”*

Also, according to the standards: “In evaluation or accountability settings, test results should be used in conjunction with information from other sources when the use of the additional information contributes to the validity of the overall interpretation.”*

The state provided growth model does not meet the standards

The state provided growth model performs differently than all other educator effectiveness indicators included in the APPR system. The other measures in the system are all based on fixed scales, the system includes 20 points based on growth or SLO (the scaling of each were described earlier), 20 points based on student achievement or growth on locally selected assessments (fixed scale where all teachers can reach effective if expectations are met) and 60 points based on evidence of teaching practice as measured by an administrator using a state approved rubric (fixed scale where all teachers can reach effective if standards are met). An analysis of the 2012-2013 APPR ratings data (*see attached APPR Distribution Table*) indicates that the results of the locally selected assessments, evidence of teaching practice and the estimated SLO score distributions were similar state-wide and by district need category. The distribution for 4-8 ELA and Math teachers that received state provided growth scores was markedly

different from the other measures since effectiveness ratings for this element are distributed relative to other teachers.

The other measures also tend to have similar distributions to each other, an indication that they are performing similarly to each other, unlike the growth model. This information supports the theory that the three other measures are a better, more stable indicator of teacher effectiveness than the single state provided growth score

Fact: The grade 3-8 New York Common Core State Assessments which are being used to produce the state provided growth measure are the same tests which the Governor and legislature say, in law (Chapter 56 part AA), should have no consequential effects on students until 2018, because of their poor implementation. They are also the same assessments which the Governor's own Safety-net legislation, that he later vetoed, was based. This law would have removed, for two years, the consequential effects of these tests on teachers receiving an ineffective or developing APPR rating.

Standard: "When tests are selected for use in evaluation or accountability settings, the ways in which the test results are intended to be used, and the consequences they are expected to promote, should be clearly described, along with cautions against inappropriate uses."*

The state provided growth model does not meet the standard

It is questionable if the grade 3-8 state assessments should be considered even appropriate gauges of student achievement much less of educator effectiveness. An estimated 55,000 students "opted-out" of the assessments during the 2013-14 school year and students that do take them may not take them seriously given that they have no consequential effects for them.

The use of these tests, if proven to be appropriate gauges of educator effectiveness, in a growth model might be justified on the grounds it will improve the quality of education by providing useful information to educators to improve instruction and promote better performance. But that is not the case in New York state's high-stakes accountability system, teachers are not allowed to score their own students assessments; test results are not released until months after the students have left the teacher's classroom; a limited number of questions are released for teachers to analyze and use for improving instruction; and, because it is based on a relative scale, it cannot be explained to a teacher what they need to do to improve their growth score.

The data also shows us that teachers' ratings based on the growth model fluctuate. If test scores were an accurate measurement of teacher effectiveness, "effective" teachers would rate high consistently from year to year because they are good teachers; and one would expect "ineffective" teachers to rate low in terms of test scores just as consistently. Instead, over two years of comparative data, approximately 33 percent of teachers moved from one category to another, and 7 percent moved two categories. The year-to-year instability this highlights suggests that test scores have very little to do with the effectiveness of a single teacher and have more to do with the assignment of students from year to year or other factors.

CONCLUSION —

Based on this analysis, clearly the New York state growth model is not meeting the Standards for Educational and Psychological Testing and should not be used as an element of the teacher evaluation system. Increasing the weight of the growth model as the Governor and Chancellor Tisch have suggested would be unfair to teachers and increase the high stakes nature of the new common core state tests.

NYSUT recommends that to improve student learning and support teacher development, the state needs to reduce the emphasis on standardized testing and expand local control of APPR through collective bargaining. Measures of student performance included in a teacher's evaluation should be determined locally. The state growth model should no longer be used in the APPR.

- **2014 American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). Standards for Educational and Psychological Testing. Washington DC: American Educational Research Association.**

APPR Distribution Table

NYS total	Ineffective	Developing	Effective	Highly Effective
Growth	8%	13%	75%	4%
Growth and SLO	4%	9%	44%	43%
SLO estimate	3%	8%	37%	53%
20% Local	1%	5%	40%	54%
60% Other	0%	2%	41%	57%
Big 4				
Growth	16%	20%	60%	4%
Growth and SLO	10%	21%	50%	19%
SLO estimate	8%	21%	47%	24%
20% Local	7%	26%	48%	20%
60% Other	0%	4%	60%	36%
Urban Suburban High Needs				
Growth	13%	16%	68%	3%
Growth and SLO	7%	13%	42%	37%
SLO estimate	5%	13%	35%	47%
20% Local	2%	12%	37%	50%
60% Other	1%	3%	46%	49%
Rural High Needs				
Growth	9%	16%	71%	4%
Growth and SLO	5%	11%	45%	39%
SLO estimate	4%	10%	39%	47%
20% Local	1%	7%	47%	44%
60% Other	0%	3%	52%	44%
Average Needs				
Growth	7%	13%	77%	4%
Growth and SLO	4%	8%	44%	44%
SLO estimate	3%	7%	36%	54%
20% Local	1%	3%	42%	54%
60% Other	1%	2%	39%	58%
Low Needs				
Growth	4%	10%	80%	6%
Growth and SLO	2%	4%	43%	51%
SLO estimate	1%	3%	35%	62%
20% Local	0%	2%	32%	66%
60% Other	0%	1%	31%	68%

Source: <https://data.nysed.gov/>. 2012-13 Annual Professional Performance Review Ratings and State Provided Growth Ratings.

Note: SLO estimate comes from subtracting the number of teachers in the State Provided Growth Rating from the combined Growth and SLO reported number.