



STRATEGIC **DATA** PROJECT

SDP FELLOWSHIP CAPSTONE REPORT

Balancing Fairness, Simplicity, and Scale in Calculating Educator Growth Scores

Joshua Marland, Regents Research Fund
Monica J. Young, New York State Education
Department

SDP Cohort 3 Fellows

CALCULATING EDUCATOR GROWTH SCORES

SDP Fellowship Capstone Reports

SDP Fellows compose capstone reports to reflect the work that they led in their education agencies during the two-year program. The reports demonstrate both the impact fellows make and the role of SDP in supporting their growth as data strategists. Additionally, they provide recommendations to their host agency and will serve as guides to other agencies, future fellows, and researchers seeking to do similar work. *The views or opinions expressed in this report are those of the authors and do not necessarily reflect the views or position of SDP or the Center for Education Policy Research at Harvard University.*

CALCULATING EDUCATOR GROWTH SCORES

Abstract

The New York State Education Department delivered educator growth scores for the first time in 2011-12 as part of its Race to the Top commitments. Doing so was a new, complex process for the State, and certainly on the largest scale nationally with approximately 38,000 educators receiving them. While the process went relatively smoothly, it was during development that there was and continues to be a very delicate balance between our primary goal for creating *fair measures* and making sound measurement decisions and two other NYSED goals, which were to include as many students and educators as possible and to make the entire process transparent and relatively easy to communicate. Through the lens of the co-project managers, this paper discusses the tensions between these goals, how NYSED engaged stakeholders to provide feedback during the process, and also considers some decisions that needed to be made by NYSED and its stakeholders so the State could provide fair measures to educators.

CALCULATING EDUCATOR GROWTH SCORES

Introduction

One of the commitments the New York State Education Department (NYSED) made when applying for the United States Education Department's (USED's) Race to the Top funding initiative, was a commitment to implement a multiple measure evaluation system for **all** classroom teachers and building principals that would ultimately factor into employment decisions for the State's approximately 250,000 teachers and principals. The three main components of this multi-measure evaluation system are:

1. Measures of improvements in student learning growth on the State assessment (or other comparable measures of growth if no State assessment is given);
2. Locally-selected measures of student achievement or growth; and
3. Other evidence of teacher or principal effectiveness as measured primarily through observation of practice in classrooms or schools.

With some high-level policy guidance and decision-making provided by NYSED, districts were primarily responsible for collectively bargaining around the allowable options and for calculating the second and third components of the evaluation system. Responsibility for calculating the first component, a student growth measure based on the State assessments (i.e., calculating a growth score) for teachers and principals in grades 4-8, ELA and Math, fell to NYSED. Like many states that either won Race to the Top dollars or have since been granted ESEA waivers, NYS had to develop and implement this substantial new initiative, and make policy and measurement decisions based on the limited large-scale research available, while coordinating these decisions across departments and stakeholder groups. To ensure this significant work would happen, the Commissioner and senior leaders set a clear vision for the work early in the process. This vision was to provide **fair** measures for educators that held them accountable for improvements in student learning and were also viewed and understood to be accurate reflections of their contribution to student learning.

In 2011-12, NYSED was required by law to calculate and issue growth scores to teachers with students in Grades 4-8 math and English language arts (ELA) and their principals. Districts that had reached a collectively-bargained new evaluation system consistent with the new Education Law §3012-c were required by law to use the State-calculated growth score data as one component of their evaluation plan for the 2011-12 school year (see early paragraph for required components to the evaluation plans). Recognizing that this State growth measure would only be calculated for about 20 percent of educators statewide, the legislation allowed NYSED to develop and add student growth models that covered additional grades and subjects in future years. Use of these additional growth

CALCULATING EDUCATOR GROWTH SCORES

models would be approved only if beta-modeling of State assessment results supported these calculations and the Board of Regents, NYSED's policy-making body, and stakeholder groups supported the use of these models on empirical and policy grounds.

As NYSED began this work, we focused on achievement of our primary goal for creating fair measures and making sound measurement decisions. This goal was achieved, but we recognize that there was some tension in that we weren't able to provide all educators statewide with State growth scores, thus requiring local districts to create comparable growth scores for these educators, and the growth model we developed was complex and not as easy to communicate in a transparent way. Some measurement decisions resulted in systematic exclusion of higher need students in the model, while others meant that fewer educators would receive growth scores, and still others made explaining how the growth scores were calculated more difficult. It was our charge as project managers to ensure that educators received fair growth scores, but we also had to identify ways that these other two goals could be met.

Growth Project Team

Monica Young, Project Coordinator at NYSED and Strategic Data Project (SDP) Agency Fellow, and Joshua Marland, Fellow at the Regents Research Fund (RRF) and SDP Data Fellow, served as the co-project managers of the development of the State growth models from inception of the request for proposals (RFP) in early 2011, through the second year of distribution of growth scores to educators in 2013. The primary responsibilities for project management fell into two main categories: Monica primarily managed development of communications and reporting, as well as all contract-related matters, while Joshua managed coordination of data requests from SED and oversaw decisions with the vendor about data and analytics that were used in decision-making and in calculation of growth scores. They worked primarily under the direction of Amy McIntosh, Senior Fellow at RRF, who was responsible for the inception and implementation of this work and Ira Schwartz, Assistant Commissioner of Accountability, who was responsible for consistency between NYSED institutional accountability measures and educator accountability measures. Finally, Ken Wagner, Deputy Commissioner for Curriculum, Assessment and Education Technology, and the Information and Reporting Services team under his direction ensured that any decisions made related to growth score calculations could be feasibly implemented, and that the data elements used were suitable for calculations. The co-project managers, Ms. McIntosh, and Mr. Schwartz are jointly referred to as "the growth project team" for the rest of this report. In addition, other senior leaders at NYSED were regularly involved throughout the

CALCULATING EDUCATOR GROWTH SCORES

process, including many of the Assistant and Deputy Commissioners, as well as the Commissioner and Board of Regents.

Project Engagement

The legislative mandate requiring a multi-measure evaluation system was just the first step of many that needed to be taken in order to develop an educator growth score methodology, but it did set the foundation for requirement of State-calculated educator growth scores. When Race to the Top funding was subsequently awarded, NYSED immediately established and engaged a Board of Regents Task Force on Teacher and Leader Effectiveness (“Task Force”) that was comprised of representatives from across the State of different stakeholder groups at various levels within the educational community, including teachers, principals, district administrators, higher education faculty, constituent group leaders, union representatives, etc. The Task Force, and a set of small working groups, met regularly to discuss and provide feedback and suggestions on NYSED’s plans for developing the statewide evaluation system. One of these working groups was called the Metrics Workgroup and this smaller group was charged with supporting the development of the State growth/value-added models. An important clarification is required regarding how NYS law defined a “growth model” versus a “value-added model”. The law defined a “growth model” as a student growth percentile model which may include consideration of student poverty, English language learner, or disability status and represents 20% of an educator’s evaluation and a “value-added model” as a model that can include additional control variables, cannot be implemented before the 2012-13 school year, must be approved by the Board of Regents, and represents 25% of an educator’s evaluation. Though most researchers would consider this definition of “growth model” to be a form of a “value-added model”, in NYS a “value-added model” has a specific regulatory meaning; thus this paper uses the term “growth model” as defined above (additional information about the differences between these models is explained in the *Project Implementation* section below).

Metrics Workgroup members had varying degrees of working knowledge with respect to the policy and statistical decisions required to develop growth models. Some had no knowledge at all, some had a sense of the “Colorado growth model” (the model developed by the Colorado Department of Education with the National Center for Improvement of Educational Assessment), whereas others had a deeper understanding because their districts were already implementing a growth/value-added model locally for their own evaluation systems and learning, or in the case of higher education faculty members, they were involved in research on growth/value-added models. The growth project team’s

CALCULATING EDUCATOR GROWTH SCORES

charge was to bring all of the Metrics Workgroup members to a base level of understanding and familiarity so that they could develop a high level of comfort with the growth model in advance of NYSED's distribution of growth scores in 2011-12, and so they would ultimately support the recommendation to the Board of Regents to adopt the use of a "value-added model" in future years.

On a fairly regular basis, the growth project team along with the vendor who was selected to produce the growth and value-added models, the American Institutes for Research (AIR), would present topics for discussion to the Metrics Workgroup such as: how NYSED would use district-provided data to link students to educators; whether a minimum time should be required for an educator/student linkage to "count" for growth score purposes; and, how many student scores were required for an educator growth score to be calculated or reported (these examples are expanded on later in the report). After the initial growth model was developed for use in 2011-12, the Metrics Workgroup continued to provide feedback on additional topics such as: whether or not NYSED should enhance the Grades 4-8 ELA and math growth model in 2012-13 to include additional student demographic information; how NYSED could refine the way students were linked to educators; and, how State assessment data could be used to develop a new model for high school principals.

In addition to engaging with the Metrics Workgroup, the growth project team also met regularly with NYSED's Information and Reporting Services (IRS) team to discuss all issues related to the data that would ultimately get used in State growth calculations. Since all but one member of the growth project team was relatively new to NYSED and RRF, the IRS team played a significant role in providing background information about the quality of data elements, the timing for data collection, and nuances within each data element that could impact any inferences drawn from the analyses. Finally, the IRS team was also ultimately responsible for providing the vendor with the data used for the growth analyses that were collected from Districts, the student assessment data, and other SED data and then for storing the results data produced from the growth analyses. Over time, the growth project team became responsible for providing the background information on the growth analyses to ensure that the IRS team understood the results of the growth analyses that they would be storing.

Other groups were also consulted during development and implementation of the growth model on a more ad-hoc basis to ensure that other perspectives were included and that the State was meeting its regulatory/legislative requirements for developing these measures. The growth project team presented and discussed the specifics of the model with: NYSED's growth model technical advisory committee which was comprised of prominent researchers with expertise in growth and value-added methodologies; the Strategic Data Project fellows and leaders at the Cohort 3 Memphis workshop and at

CALCULATING EDUCATOR GROWTH SCORES

the 2013 Annual Convening; cross-state education agency consultancies; and, a statewide organization of practitioners specifically devoted to NYS data. A senior leader project team (that included other NYSED Deputies and Assistant Commissioners) also met regularly with the growth project team to make decisions based on the feedback gathered through these different working groups and to come to consensus on various issues for presentation to the Commissioner and the Board of Regents.

Project Implementation

The project team envisioned the work as two distinct phases – the “growth model” year in 2011-12 where student’s prior academic achievement, disability status, English language learner status and poverty status all could be included as covariates (see *Year One: Growth Model* section below for details on covariates). After that year, NYSED could add covariates and increase the weight of growth scores in educators’ evaluations if empirical and policy considerations merited it. As part of Education Law §3012-c, the addition of covariates beyond the original four was considered as the distinction between a “growth model” and a “value-added model.” The legislation prohibited using a value-added model prior to 2012-13, so the team had at least one year to work with AIR to provide a high-quality growth model, distribute the growth scores, and provide documentation and training materials prior to moving to a new value-added model.

Year One: The Growth Model

The “growth model” year was the initial phase of the work that included the “from the ground up” development. This included a request for proposals, selection of a vendor, development of the model, and delivery of growth scores statewide during the summer of 2012. This first phase of work, more so than the second, was guided by statute which, as mentioned, limited the scope of the covariates used in the growth model to four covariates. It also specified that a growth percentile be used as the reporting metric for the growth model, and it limited the weight of growth scores to 20 percent of an educator’s evaluation for 2011-12. Beyond that, the statute provided few specifications for developing the production model that would be used to calculate growth scores for release in the summer of 2012, leaving much work to be done.

The growth project team had many decisions to wrestle with during the growth model year, most of which fall into seven buckets as noted in Table 1 below. (See appendix for more specific questions along with the decisions made during each year.)

Table 1: Primary Focus Areas for Year 1

CALCULATING EDUCATOR GROWTH SCORES

Decision Category	Decision Consideration
Model Choice	Determine what type of statistical model would be used, whether a model should be developed specifically for NYS, and what measures would be reported
Covariates	Consider the quality of the data elements that were available for use as covariates in the model, and which of the elements would be used
Assessment Considerations	Working within the bounds of the State assessments, and recognizing that there would be new assessments in 2012-13, determine how test measurement error should factor into the model
Minimum N Size	Decide whether an educator's growth score should be based on a minimum number of students or a minimum number of student scores
Attribution	Determine whether and how students who were linked to a teacher or principal for less than a full year should be counted in the educator's growth score
Classification	Using the effectiveness rating categories defined by statute (highly effective, effective, developing, ineffective) and associated text descriptions, determine how educator growth scores would be used to determine an educator's rating category and rating points
Reporting	Determine what data, information, and training would be reported, to whom, by whom, and when

Year 1 Decision: One example of the considerations during year 1 was the decision for a minimum number of student scores for calculating a teacher growth score. The team considered various minimum N sizes of students and how that decision impacted the coverage of educators who would ultimately receive a growth score. NYSED could have provided a growth score to any educator with one "valid" student who was enrolled for any amount of time with the educator, thereby providing growth scores to as many educators as possible, but best practice (and common sense) would advise against that. On the flip side to that, raising the minimum N by one student per teacher would ultimately exclude ALL students in classrooms where that is a meaningful difference. For instance, a change from 18 to 19 student scores might yield a marginal improvement in the stability of the estimate, but any teacher with 18 students now gets excluded from receiving growth scores, and no teacher is ultimately accountable for the academic growth of those students within this measure.

CALCULATING EDUCATOR GROWTH SCORES

Year Two: Additional Covariates

The second phase (year two) was when NYSED was allowed by statute to move to a “value-added model,” which would include additional covariates and carry with it additional weight in the overall evaluation system. In order for that to happen, the Board of Regents would need to approve the use of a “value-added model.” Learning from the work accomplished during the first year of implementation, the growth project team developed a shorter list of focus areas that needed to be addressed in year two, listed in Table 2 below.

Table 2: Primary Focus Areas in Year 2

Decision Category	Decision Considerations
Covariates	Determine the utility of adding covariates into the model and whether adequate, reliable data are available to do so
Attribution	Consider the additional data elements collected under the TSDL (Teacher-Student Data Linkage) initiative and whether adequate, reliable data are available to enhance the way that students are linked to educators
Additional Models	Consider the State assessment data that are available and how models can be developed for use with high school principals that adequately represent the principal’s role in increasing student learning growth
Weighting Models	Consider how multiple growth scores that cover a particular educator’s assignment area can be combined to result in a single growth score

Year 2 Decision: One example from year 2 was the decision to utilize a more sophisticated approach to attribution of students to teachers. In year 1, students were attributed to teachers if they were continuously enrolled in that teacher’s classroom for 195 calendar days in English language arts (ELA) or 203 calendar days in math. The State knew when it made that decision in year 1 that some stakeholders would think it was too restrictive because requiring 195 or 203 days would exclude many high-need students who are often more mobile (moving regularly during the school year). At the same time, we heard from teacher representatives that requiring a shorter timeframe of enrollment could be unfair to teachers because the student had not received a full year of instruction from the same person.

Given that, NYSED considered various minimum enrollment durations (20% through 80%) as a cutoff for a student to be included in the calculations, and even went one step further and considered the use of current year attendance as a weight. The growth project team presented options to the

CALCULATING EDUCATOR GROWTH SCORES

Metrics Workgroup on multiple occasions, which generated difficult and vigorous conversations among the stakeholders. Metrics Workgroup members felt strongly that educators should be held responsible for all students, but they did not want to unfairly penalize a teacher for a student who switched classes (or schools) multiple times a year.

While analyzing the data, the State learned that a minimum enrollment duration above approximately 80% excluded students at a similar rate to the year prior (which seemed reasonable). The State also learned that including attendance as a weight in growth score calculations did not really change a teacher's growth score, presumably because the impact of low (or high) attendance is likely reflected in a student's test score and also likely does not change much year over year. Taking this information into consideration, the State committed, in the second year of this work, to an approach that required a lower minimum enrollment duration of 60%, but also included each student's attendance duration in the course as a weight.

Many of the conversations the growth project team had during year one and year two as we attempted to answer these questions required the use of data and careful consideration of our own ability to implement and communicate these decisions, as well as the potential implications for educators specifically and the educational community more generally. These conversations required discussion between AIR, senior leadership, and the growth project team, and all were presented to the Metrics Workgroup and Task Force before the decision was made for use in either year.

Challenges to Implementation

In considering the options related to many of the required decisions above, the growth project team had to attend to several challenges during development. As mentioned, the first was a tension between the statistical precision of a measure and the breadth of students and/or educators included in the measure. The decision to require that an educator have 16 student scores attributed to him or her was based on data about the reliability of the growth measure as you decrease the number of student scores included in the measure, and the decision to require that a student be linked to a teacher for at least 60% of the course duration was based on analytics showing the differences between attribution approaches. The challenge became finding a solution that preserved the reliability of the measures, ensured that educators were being held accountable for our highest-need students, and ensured that many of our 4-8 ELA and math teachers and principals would receive State-provided growth scores.

Another challenge was what we saw as the inverse relationship between simplicity and fairness in growth models. NYSED could have chosen a model that was relatively easy to message to

CALCULATING EDUCATOR GROWTH SCORES

stakeholders and educators, but the analytics failed to provide sufficient evidence that the simplest model would account for all the complexities that existed in classrooms and schools across New York State. NYSED felt that these complexities could only be accounted for with a more sophisticated model, which resulted in a more statistically complex model, in order to include many of the covariates the Metrics Workgroup members suggested, and to retain the fairly sophisticated attribution and classification approaches.

The growth project team was also keenly aware of the level of patience, understanding, and constant solutions-orientation that was required to implement such a complex process on this large of a scale. The data collection process, in and of itself, was a difficult one – NYSED needed to collect TSDL information for the first time from every public school district in New York – a data file that when aggregated statewide ultimately surpassed 10 million records for the 2011-12 school year (and 40 million records in 2012-13). The State, and in particular the IRS team, spent considerable time in conversations with districts and their data centers as the field tried to comprehend and implement the requirements for data collection since this was a new process and NYSED’s high-level guidance could have been interpreted differently.

In addition, the business rules and data collection timelines for such a complex process required a deep level of understanding across data systems and people; thus, IRS’s constant involvement was paramount for success in this project. Another issue NYSED had to compete with was the assessment administration and scoring timelines and the statutory requirement for providing State growth scores to districts in time for inclusion in an educator’s annual evaluation results (evaluations must be finalized by September 1st of each year). These timelines gave AIR a razor-thin number of weeks from data intake until completion of calculations and reporting of results to the State and educators.

Another challenge, which still remains, is in explaining the complex statistical modeling used to calculate State growth scores to NYSED staff members, who are incredibly knowledgeable about data systems, data management, or the teaching profession, but are not as knowledgeable about how these come together in creating a statistical model for educator effectiveness. Conversely, staff members in these other offices needed to educate the growth project team and AIR on the complexities of teacher and principal data collection at the State level.

Related to that, effectively communicating the differences between growth and status measures is really difficult. Many within the educational community would argue for the use of a growth measure because of the issues related to status achievement measures, but aspects of the concept of growth on non-vertically scaled assessments (and similar student comparisons) are often lost in translation. This

CALCULATING EDUCATOR GROWTH SCORES

could also be a product of communicating more generally, in that the really complex calculations require a lot of time to communicate, and even longer for stakeholders to digest and understand.

Outcomes

NYSED, in collaboration with AIR, developed and delivered growth scores to educators in August 2012, and implemented and launched a growth reporting system in November 2012, thus meeting the goal set forth in the statute. The NYSED growth project team immediately began work on phase two of the project: development of a value-added model and inclusion of additional grades and subject areas in the model for implementation in future years. Ultimately, the Board of Regents voted to use an enhanced version of our 2011-12 growth model with additional data around the four “growth model” covariates (an “enhanced growth model”) for the 2012-13 and 2013-14 school years while approving the transition to a “value-added model” for 2014-15, at which point it will represent 25 percent of an educator’s evaluation.

The growth model development and on-going enhancement work is a fundamental aspect of implementation of the Regents Reform Agenda, specifically the need to have a statewide annual evaluation system. As such, NYSED will continue to engage stakeholders in refining the model in future years. In particular, NYSED will continue to ensure that the State-provided growth scores hold educators accountable for all students’ academic growth. Historically, high-need students tend to get lost in accountability systems, and NYSED has, and will continue to, actively seek ways to be sure these students are attributable and accountable to educators.

Lessons Learned

The SDP Fellows identified a number of areas where we gained knowledge, experience, and skills over the course of implementing this project and as a result of our involvement with SDP. With respect to the project discussed in this report, our lessons learned are summarized by topic as noted below.

Engagement. The facilitation of stakeholder meetings to gain meaningful and actionable feedback was challenging, as it required ensuring everyone involved had a base level of understanding so they felt comfortable providing feedback. Thus the process of engagement required several meetings where the growth team, AIR, and outside researchers presented background information and research to help stakeholders attain this base level of knowledge so that they would be prepared to consider the

CALCULATING EDUCATOR GROWTH SCORES

models AIR developed. Incorporating the stakeholders' feedback into the evaluation system also required careful consideration by the State as the stakeholders often represented diverse and sometimes opposing perspectives from each other and from the State.

Content Expertise. It was (and still is) essential to have a deep understanding of both the inputs (i.e., the data elements that are used in a model) and the outputs (i.e., how to interpret the results of the analyses). In addition, making decisions based on the outputs required knowledge of the political environment and assessment of impact, thus this project team needed to have diversity of content expertise to bridge that gap. It was also essential to have extensive discussion and input from leaders who have broad perspective on the reform agenda of the State as well as the historical background on these initiatives.

Tradeoffs. As mentioned throughout this report, we felt that there is an inverse relationship between simplicity and fairness. A simple model that could be easily communicated may unfairly represent the academic growth of students, or may unfairly represent the role of teachers or principals in helping the students achieve academic growth, whereas a fair model could be extremely difficult to communicate because of the advanced statistical basis.

Development Time. It is imperative that the amount of time it takes to make data and analyses fully operational for the field not be underestimated. This is particularly important when considering how to collect and use new data elements and when considering the use of sophisticated calculations. Beta-modeling with prior year data provided a host of information to develop a model, but rules became more complex as we operationalized ran the calculations using data from the current year.

Conclusion

We learned a great deal over the course of the past two and a half years regarding what it takes to develop and implement a large-scale data calculation process that has high stakes attached to it. More than anything, we realized there is a genuine need for people with quantitative experience like ourselves, who are able to bridge the gaps between data collectors who manage data and policymakers who use those data to make decisions. Without the commitment of all three groups a very different outcome could have been yielded for the State. Our success is evident in that we were able to deliver a second year of results to educators this summer and have plans to do so for a third year in summer 2014.

CALCULATING EDUCATOR GROWTH SCORES

Appendix 1: Questions addressed during years one and two

Year One: 2011-12

Decision Category	Considerations	Decision
Model Choice	Should NYSED adopt the “Colorado growth model”, or some other covariate-adjustment model? How difficult would it be to use a covariate-adjusted model and assign a growth percentile as the reporting metric?	Mixed model covariate adjustment model that includes the use of demographic variables
Covariates	How many years of prior achievement should NYSED include as covariates? Should NYSED use demographic controls when calculating growth?	Up to three years for students
Assessment Considerations	Should NYSED acknowledge there is measurement error in State test scores and attempt to mitigate its impact on growth calculations?	Account for measurement error in both the predictor and outcome variables
Minimum N Size	What should the minimum N size be for calculating a teacher or principal’s growth score? How much reliability is gained at various minimum N sizes? What is the tradeoff between ensuring reliable measures and providing growth scores to as many educators as possible?	16 student scores across ELA or math (this could be 8 students who all have ELA and math or 16 students who only have ELA)
Attribution	How should NYSED account for students who spend less than a full academic year in the classroom?	Students must be continuously enrolled for 195 calendar days in ELA or 203 days in math
Classification	How should NYSED classify teachers and principals into the four levels of effectiveness (highly effective, effective, developing, ineffective)? How can NYSED use confidence intervals in its classification approach? How should the 20 points available for the growth score be assigned to educators?	Use standard deviations and confidence intervals for classification. For instance, teachers whose scores are 1.5 standard deviations above the mean and whose lower bound CI does not cross over the State average are considered Highly Effective
Reporting	What should a teacher or principal’s growth report look like? What data should be included in the report? What resources will educators need to interpret their growth reports? How will NYSED deliver growth reports to educators? How will NYSED provide training and technical assistance regarding use and interpretation of growth scores?	Turnkey training at statewide network team institute; tutorial videos; annotated growth reports for teachers and principals. See EngageNY.org

CALCULATING EDUCATOR GROWTH SCORES

Year Two: 2012-13

Decision Category	Considerations	Decision
Model Choice	Should NYSED adopt the “Colorado growth model”, or some other covariate-adjustment model? How difficult would it be to use a covariate-adjusted model and assign a growth percentile as the reporting metric?	Mixed model covariate adjustment model that includes the use of demographic variables
Covariates	What additional covariates can NYSED add to the model to ensure fairness? Which demographic characteristics have the most impact on the models? Do any of the covariates have potentially negative political or face validity implications?	Enhanced growth model covariates, including range of prior achievement in class, prior-year NYSESLAT score (language proficiency assessment), spends < 40% of time in general education setting, prior year achievement in other subject
Attribution	How can NYSED enhance its attribution approach to account for enrollment duration in a class? Should NYSED consider including attendance as a factor in calculating growth scores? How can NYSED use section codes to better attribute students to teachers and better define classroom characteristics?	Students must be enrolled in a teacher’s class for 60% of the year, and will then be weighted by their attendance in a teacher’s growth score calculations
High School Models	How can NYSED account for the improvements in learning at the high school level that principals are accountable for? What are the typical Regents-taking patterns of a NYS high school student? How does Regents-taking vary across districts? Using currently collected data, can individual principals be differentiated from schools? How should multiple measures be combined to create one growth score?	Two models were adopted for principals: 1) HS mean growth percentile for English and Algebra I Regents exams 2) Growth in Regents Exams Passed, which measures the extent to which principals are able to have their students pass more Regents exams than similar students statewide
Reporting	How can NYSED combine growth scores across high school models AND across grade levels (4-8 and HS) to create one growth score for a principal?	Student weighted for each level