Teacher and Principal Evaluation Improvement Workgroup Summary Report

MARYLAND STATE DEPARTMENT OF EDUCATION
EQUITY AND EXCELLENCE

Office of Leadership Development and School Improvement

March 2019

Maryland State Department of Education

Office of Leadership Development and School Improvement
Tiara Booker-Dwyer, Executive Director
Edmund Mitzel, Coordinator of Leadership Development
Dr. Brian Eyer, Leadership Development Specialist
Dr. Mary Minter, School Leadership and Turnaround Support Specialist
Dr. Felicia Lanham Tarason, School Leadership and Turnaround Support Specialist
Laura Liccione, School Leadership and Turnaround Support Specialist
Dr. Li Wang, Data Analyst for Educator Effectiveness and School Improvement
Tara Corona, Continuous Improvement Specialist
Teri Windley, Management Associate
Hazel Hamond-Terry, Management Associate

TABLE OF CONTENTS

Evaluation Improvement Workgroup Members

The Office of Leadership Development and School Improvement would like to thank everyone who participated in the Evaluation Improvement Workgroup. Workgroup sessions were facilitated by the Mid-Atlantic Comprehensive Center @West Ed. Research was provided by the Regional Educational Laboratories Mid-Atlantic (REL Mid-Atlantic) at Mathematica Policy Research.

Name	Organization
Terry Ball	Maryland Association of Elementary School Principals
Tiara Booker-Dwyer	Maryland State Department of Education
Cheryl Bost	Maryland State Education Association
Dr. Faith Connolly	Baltimore Education Research Consortium
Latisha Corey	Maryland Parent Teacher Association
Juliann Dibble	Howard County Public Schools
Dr. Matthew Duque	Maryland State Department of Education
Dr. Bonnie Ennis	Wicomico County Public Schools
Dr. Brian Eyer	Maryland State Department of Education
Tom Evans	Maryland Association of Secondary School Principals
Dr. Linda Gill	Pubic School Superintendents' Association of Maryland
Dr. Steven Glazerman	REL Mid-Atlantic at Mathematica Policy Research
Dr. Laurie Henry	Salisbury University, School of Education
Lisa Hopkins	Caroline County Public Schools
Dr. Rhonda Jeter	Bowie State University, College of Education
Laura Liccione	Maryland State Department of Education
Carla McCoy	Baltimore Teachers Union
Angela Minnici	Mid-Atlantic Comprehensive Center @ WestEd
Edmund Mitzel	Maryland State Department of Education
Dr. Christopher Morphew	Johns Hopkins University, School of Education
Janet Pauls	Queen Anne's County Public Schools
Nancy Reynolds	Maryland Association Boards of Education
Dr. Kim Rotruck	Frostburg State University, College of Education
Rudy Ruiz	Maryland Business Roundtable for Education
Dr. Carolyn Teigland	Public School Superintendents' Association of Maryland
Jeffrey Terziev	REL Mid-Atlantic at Mathematica Policy Research
Dr. Elias Walsh	REL Mid-Atlantic at Mathematica Policy Research
Dr. Mary Young	Prince George's County Public Schools

Overview and Background

The Maryland State Department of Education is committed to ensuring that each student has effective teachers and educational leaders. Essential to this commitment are valid and reliable evaluation systems. Evaluation systems that accurately differentiate performance levels can inform support required to enhance the professional practice of teachers and school leaders and improve educational outcomes for students.

In 2010, a law (Education Reform Act of 2010) was enacted that requires the Maryland State Board of Education to adopt regulations that, "….*establish general standards for performance evaluations for certificated teachers and principals that include observations, clear standards, rigor, and claims of evidence of observed instruction*." In 2012, the Maryland State Board of Education adopted regulations to guide the evaluation of teachers and principals. The regulations consisted of requirements for a state default model (Code of Maryland Regulation 13a.07.09.05) and a local education agency evaluation system (Code of Maryland Regulation 13a.07.09.04). The state evaluation model consists of equally weighted measures of professional practice and student growth. The professional practice domains for teachers are planning and preparation, classroom environment, instruction, and professional responsibilities. The professional practice domains for principals were the Maryland Instructional Leadership Framework and Educational Leadership Policy Standards.

The adopted regulations established a foundation for the evaluation of teachers and principals in Maryland. However, these regulations failed to define standards, rigor, and claims of evidence of observed instruction as required by state law for the evaluation of teachers. Student learning objectives (SLOs) are the primary measure of student growth for teachers and principals in the adopted regulations. There are several benefits to using SLOs as growth measures. SLOs can be used with all teachers and they allow teachers to establish their own goals based on the needs of their students. There is increasing research that suggest SLOs may not provide a valid or accurate measure of a teacher's contribution to student learning. The implementation of adopted regulations over the last four years resulted in over 95% of teachers and principals being rated effective or highly effective. The inflated evaluation results coupled with the decline in student performance on state assessments drew concern around the reliability and validity of Maryland's evaluation system.

In February 2017, the Maryland State Board of Education adopted the Professional Standards for Educational Leaders (PSEL). The PSEL defines the practice of an effective school leader to support the academic success and well-being of each student. These standards replaced the Maryland Instructional Leadership Framework and Educational Leadership Policy Standards. The newly adopted PSEL guide administrator preparation, licensure, and evaluation in Maryland. In July 2018, the Office of Leadership Development and School Improvement in collaboration with the Community Training and Assistance Center (CTAC) and stakeholders developed a PSEL rubric. The Maryland PSEL Rubric builds off the practices identified for an effective leader in the PSEL document by expanding the definition to include practices of highly effective, developing, and ineffective administrators. In October 2018, the Office of Leadership Development and School Improvement released a draft Principal Evaluation Guidebook. This guidebook was designed to support principal supervisors in facilitating the evaluation process. It offers guidance, strategies, templates, and sample evidence that will support effective evaluation practices. The PSEL rubric and guidebook serve as a foundation for principal evaluations in Maryland. The Office of Leadership Development and School Improvement provides training on the PSEL rubric to support effective evaluation practices and improve inter-rater reliability.

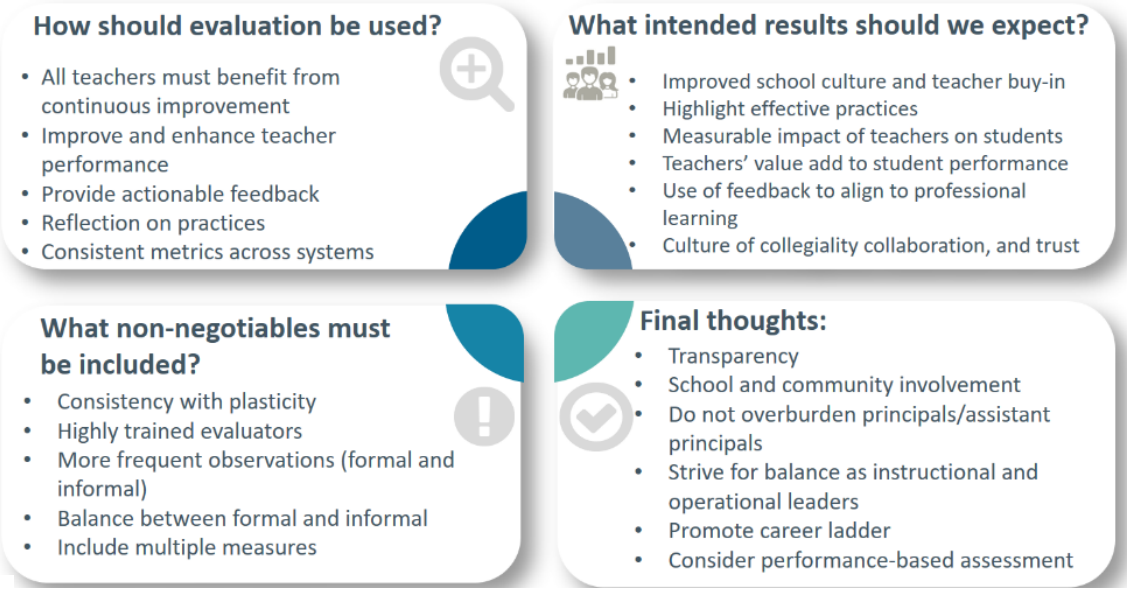
In September 2018, an Evaluation Improvement Workgroup was convened to inform improvements to the evaluation system focusing on the professional practice domains for teachers and student growth measures for teachers and principals. The workgroup was required to establish recommendations that were grounded in research; informed by data; focused on elevating professional practice and improving student performance; and complied with the requirements established in the Education Reform Act of 2010. The recommendations developed by the workgroup will be used to inform revisions to the teacher and principal evaluation system.

This report summarizes evaluation models explored and identifies preliminary recommendations of the workgroup. The recommendations of the workgroup will be shared with stakeholders (representatives from school, school systems, institutions of higher education, etc.) for additional input prior to revising the regulation for the state model.

Revising the Evaluation System

The workgroup began the process for making recommendations to revise the evaluation system by mapping the desired state for teacher evaluations. Figure 1 summarizes the feedback from workgroup members regarding the desired state.

Figure 1: Desired State for the Teacher Evaluation Model



Teacher Professional Practice

The Evaluation Improvement Workgroup (workgroup) analyzed frameworks that were commonly used in states for the evaluation of teachers. The workgroup reviewed the following frameworks to inform recommendations for the teacher professional practice domains:

- The Framework for Teaching (Charlotte Danielson Four Domains Model)
- The Framework for Teaching Clusters (Charlotte Danielson)
- Classroom Assessment Scoring System (CLASS)
- The System for Teacher Advancement (TAP)
- The Marzano Teacher Evaluation Model

Workgroup members received an overview of the framework, summary of research that supported the framework, and a list of states that currently implement the framework. Members of the workgroup were asked to respond to the following questions for each framework:

1. What resonates with you?
2. How well does this framework capture what teachers should know and be able to do?
3. Are there any gaps?
4. How well does the framework align with the desired state for evaluations?

Each workgroup member was asked to share information about each framework with their constituents to identify a preferred framework. Workgroup members shared feedback at the October 2018 meeting. Each member reported back to the whole group the thoughts and concerns from their representative organizations (refer to Appendix I for summary of feedback from each workgroup meeting).

There was preliminary consensus to adopt the Danielson Frameworks. There are currently two versions of the Framework for Teaching – the four domains version and the cluster version. There was not a consensus on which version of the framework to adopt.

The rationale for the selection of the Danielson Frameworks was as follows:

1. The Danielson Framework for Teaching is grounded in research and aligned to the Interstate New Teacher Assessment and Support Consortium.
2. The Danielson Framework for Teaching is informed by data which resulted in the updated cluster version.
3. The Danielson Framework for Teaching is rigorous and includes a rubric with four categories of performance (Distinguished, Proficient, Basic, and Unsatisfactory). The clusters version includes rubrics specific to English language arts and mathematics instruction.
4. 21 of the 24 school systems in Maryland are currently using the Danielson Framework or a modified version of the Danielson Framework.

Recommendation for Teacher Professional Practice

It is being recommended that the state evaluation model include the Danielson Frameworks - four domains and cluster versions - and associated rubrics as measures for teacher professional practice domains.

Student Growth Measures

The Evaluation Improvement Workgroup was charged to identify student growth measures for teachers and principals. Representatives from the Regional Educational Laboratory at Mathematica Policy Research provided research and evidence of effectiveness for four different student growth measures.

1. Student Learning Objectives (SLOs) - Measure whether educators met established student learning goals for their students.
2. Student Growth Percentiles (SGPs) – Measure how well students progress on outcomes compared with progress of other students who performed similar.
3. Educator Impact – Measures how well students progress on outcomes compared with progress of similar students and accounts for factors outside educators' control (attendance, prior assessment scores in other subjects, etc.).
4. School Wide Measure – Measure connected to overall school performance rating.

A fifth measure was also discussed where local school systems created assessments for each subject area for use in evaluation. The Prince George's County Public Schools workgroup member shared how this is used in her school system.

For each growth measure, the following characteristics were considered:

- Coverage – what share of teachers can use the growth measure?
- Effort – what is required to create the growth measure?
- Fairness – is the measure correlated to student characteristics?
- Simplicity – how easy is it to understand/explain the measure?
- Validity – does it accurately measure teaching effectiveness?
- Reliability – how consistent is the measure across time?
- Timeliness – are growth measures current data or lagging data?
- Standardization – how objective is the measure across different schools and school systems?

Figure 2 summarizes characteristics for each growth measure discussed in the workgroup.

Figure 2: Growth Measure Characteristics

	Student Learning Objectives	Student Growth Percentiles	Educator Impact	School-wide
Coverage	High	Low	Low	High
Effort	High	Low	Low	Low
Fairness	Low	Med	High	Low
Simplicity	Low	Med	Low	Med
Validity	Low	Med	High	Low
Reliability	?	Med	Med	High
Timeliness	High	Low	Low	Low
Standardization	Low	High	High	High

Workgroup members also reviewed strengths and limitations for each growth measure. The review was provided by representatives from the Regional Educational Laboratory at Mathematica Policy Research

and was based on the publication, "Measuring Progress in the Classroom: How do Different Student Growth Measures Compare? (Fact Sheet)" Figure 3 summarizes strengths and limitations discussed in the workgroup.

Figure 3: Summary of Strength and Limitations for Each Growth Measure
Content in the table is adapted from: Measuring Progress in the Classroom: How do Different Student Growth Measures Compare? (Fact Sheet)"

Student Growth Measure	Strengths	Limitations
Student Learning Objectives	• Allows teachers to set their own goals, so may be viewed as more connected to instructional improvement.	• May not provide a valid or accurate measure of a teacher's contribution to student learning. • Does not use a statistically rigorous process and may not sufficiently account for factors outside of teachers' control. • Difficult to meaningfully compare performance across teachers. • Difficult to implement rigorously and consistently.
Student Growth Percentiles	• Accounts for fewer factors outside teachers' control, which may make the approach conceptually easier to understand than educator impact models.	• Less evidence of validity compared to educator impact models. • Accounts for fewer factors outside of teachers' control and could result in less accurate evaluations. • Like impact models, student growth percentile models are statistically complex and can be calculated only for teachers of grades and subjects with the requisite student test scores.
Educator Impact	• Valid measure of teachers' impacts on students.	• Limited to teachers of grades and subjects with the requisite student test scores. • Statistically complex and can be difficult to report or explain clearly to stakeholders.
District-Designed Assessments	• Every teacher in every subject would have an accountability measure. • Local control is supported.	• Time and labor intensive for each school system. • Concern regarding validity and reliability of locally made assessments.

Recommendation for Student Growth Measure

The members of the work group reviewed five options for student growth: student growth percentiles, student learning objectives, educator impact model, school-wide measure, and district made assessments. A sixth option was suggested by workgroup members. The sixth option would be to delay identifying a student growth measure until more information is known about the Maryland Comprehensive Assessment Program (MCAP). The overwhelming consensus of the work group was in favor of option six.

Student Surveys

Student surveys are a powerful resource to obtain data that can inform improvements to an educator's professional practice. Some workgroup members wanted to engage in additional discussion around incorporating student surveys as part of the evaluation system. The use of student surveys is optional for school systems. The Maryland State Department of Education will provide information about student surveys so that school systems can make informed decisions about the use of surveys for the purpose of evaluation. Appendix 2 contains information on commonly used student surveys.

APPENDIX I: MEETING MINUTES

The information below are notes taken during workgroup meetings.

September 27, 2018 Meeting Minutes

Attendees:

- Terry Ball, Maryland Association of Elementary School Principals
- Tiara Booker-Dwyer, MSDE Office of Leadership Development and School Improvement
- Dr. Faith Connolly, Baltimore Education Research Consortium
- Latisha Corey, Maryland Parent Teacher Association (PTA)
- Juliann Dibble, Howard County Public Schools
- Dr. Matthew Duque, MSDE Office of Research
- Geraldine Duval, Maryland State Education Association
- Dr. Bonnie Ennis, Wicomico County Public Schools
- Tom Evans, Maryland Association of Secondary School Principals
- Dr. Brian Eyer, MSDE Office of Leadership Development and School Improvement
- Dr. Laurie Henry, Salisbury University
- Dr. Rhonda Jeter, Bowie State University
- Laura Liccione, MSDE Office of Leadership Development and School Improvement
- Carla McCoy, Baltimore Teachers Union
- Angela Minnici, Mid-Atlantic Comprehensive Center
- Edmund Mitzel, MSDE Office of Leadership Development and School Improvement
- Dr. Christopher Morphew, Johns Hopkins University
- Nancy Reynolds, Maryland Association of Boards of Education
- Jeffery Terziev, Mathematica Policy Research
- Dr. Elias Walsh, Mathematica Policy Research
- Dr. Li Wang, MSDE Office of Leadership Development and School Improvement
- Dr. Mary Young, Prince George's County Public Schools

The Mid-Atlantic Comprehensive Center (MACC) reviewed four professional practice frameworks for the workgroup to consider. Frameworks included: Danielson Framework for Teaching (Four Domains) and Danielson Framework for Teaching- Six Clusters; CLASS Dimensions; Marzano Teacher Evaluation Model; and TAP Teaching Skills, Knowledge, and Responsibilities Performance Standards. When workgroup participants were asked to share thoughts on existing and desired state for professional practice frameworks for teacher evaluation, the following ideas were recorded:

- Adopt a state model that allows for local variation.
- Crosswalk state model with National Board Professional Teaching Standards.
- Differentiate evaluations for new versus experienced teachers.
- Locals like to negotiate what is best for their teachers.
- SLOs are overwhelming.
- Teacher preparation needs to be connected to teacher evaluation.
- Evaluations should connect to Kirwan Commission recommendations.
- Consistency is needed across state for evaluations.
- Define "teachers of record."

- Current system is subjective and not a lot of consistency.

October 25, 2018 Meeting Minutes

Attendees:

- Terry Ball, Maryland Association of Elementary School Principals
- Tiara Booker-Dwyer, MSDE Office of Leadership Development and School Improvement
- Cheryl Bost, Maryland State Education Association
- Dr. Faith Connolly, Baltimore Education Research Consortium
- Latisha Corey, Maryland PTA
- Juliann Dibble, Howard County Public Schools
- Dr. Matthew Duque, MSDE Office of Research
- Dr. Bonnie Ennis, Wicomico County Public Schools
- Tom Evans, Maryland Association of Secondary School Principals
- Dr. Brian Eyer, MSDE Office of Leadership Development and School Improvement
- Tracey Fowlkes, Baltimore Teachers Union
- Linda Gill, Public School Superintendents' Association of Maryland
- Dr. Steven Glazerman, Mathematica Policy Research
- Lisa Hopkins, Caroline County Public Schools
- Dr. Rhonda Jeter, Bowie State University
- Laura Liccione, MSDE Office of Leadership Development and School Improvement
- Angela Minnici, Mid-Atlantic Comprehensive Center
- Edmund Mitzel, MSDE Office of Leadership Development and School Improvement
- Dr. Christopher Morphew, Johns Hopkins University
- Janet Pauls, Queen Anne's County Public Schools
- Nancy Reynolds, Maryland Association of Boards of Education
- Dr. Kim Rotruck, Frostburg State University
- Rudy Ruiz, Maryland Business Roundtable for Education
- Dr. Carolyn Teigland, Public School Superintendents' Association of Maryland
- Jeffery Terziev, Mathematica Policy Research
- Dr. Elias Walsh, Mathematica Policy Research
- Dr. Li Wang, MSDE Office of Leadership Development and School Improvement
- Dr. Starlin Weaver, Salisbury University
- Dr. Mary Young, Prince George's County Public Schools

1) Workgroup participants were asked to share feedback from their organization as to which professional practices framework to adopt, adapt, or modify. Most (12 of 19) workgroup participants selected Danielson's Framework for Teaching (the 10 members from MSDE, Mid-Atlantic Comprehensive Center at West Ed, Mathematica Policy Research abstained from the vote).

   In general, thoughts on changing a professional practice framework with a new professional practices framework for teachers was concerning. Participants representing educators across the state expressed concern in making a significant change in the professional practice for teacher evaluation while adjusting to meet requirements in the Maryland Every Student Succeeds Act (ESSA) Consolidated Plan.

2) REL Mid-Atlantic at Mathematica Policy Research experts presented three student growth measures for consideration to workgroup participants. Participants examined the State's current student growth measure- student learning objectives (SLOs) – and discussed advantages and disadvantages in maintaining this measure. Research shared showed reason for re-examination of this student measure:
    - Goals differ across teachers, making comparisons difficult.
    - May not fully account for prior skills of students and/or other factors outside of teachers' control.
    - Burdensome to implement effectively.
    - No evidence of reliability and weak evidence of validity.
    - Unknown as to the connection of student learning objectives making positive impact on student performance.

3) The three student growth measures, presented by REL Mid-Atlantic, for consideration included:
    - Student Learning Objectives;
    - Student Growth Percentiles; and
    - Education Impact (value-added) model.

    For each growth measure, workgroup members discussed what resonates with them and what questions remain. Figure 4 summarizes the discussion of the workgroup.

Figure 4: Summary of Workgroup Discussion of Student Growth Measures

Student Learning Objectives		Student Growth Percentiles		Educator Impact (value-added)	
What resonates?	What questions remain?	What resonates?	What questions remain?	What resonates?	What questions remain?
▪ Leads to regular conversations ▪ SLOs should not create barriers to effective instruction ▪ Do not work as well for all student groups	▪ Is this still an option? ▪ What replaces it? ▪ How do we link student growth to SLOs?	▪ Effective use of data ▪ Limitation of model for use with teachers in state-assessed areas ▪ State captures all of the data necessary for this model ▪ Meets requirements for ESSA accountability indicator ▪ Concern over expecting local systems to purchase assessments	▪ If local school systems agree to this measure, what assessments would satisfy this requirement? ▪ How does this account for student mobility? ▪ How would this work in application for non-tested areas, principals, and teachers?	▪ Single test measure is a plus ▪ Factors in student characteristics for which teachers and principals have little control ▪ Equity measure (factors such as student attendance, mobility, race, etc.) ▪ Considers for gaps in content learning	▪ What about letter grades? ▪ How does this relate to teacher evaluation?

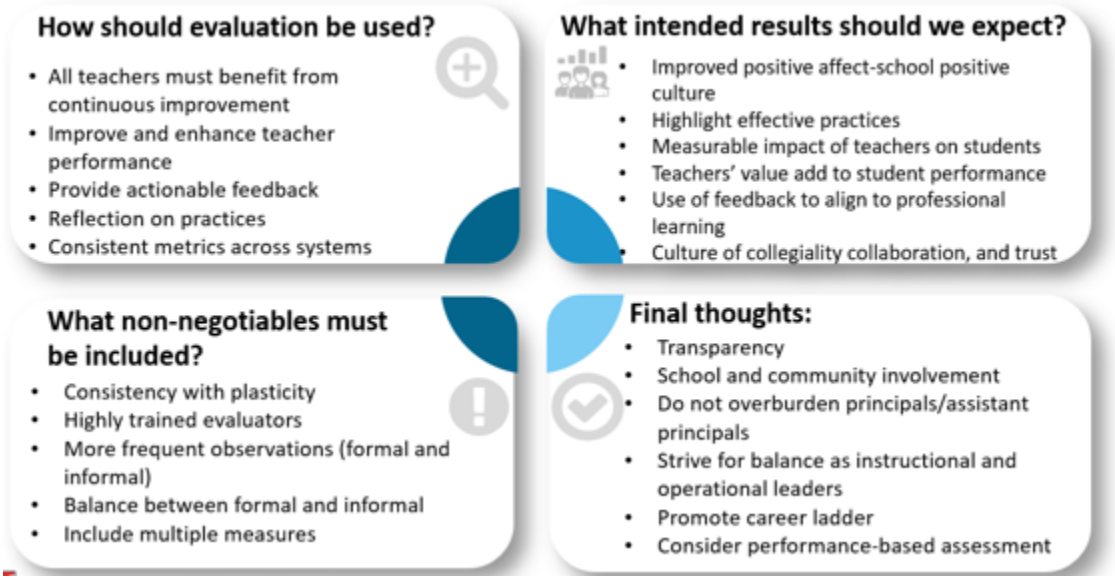
November 11, 2018 Meeting Minutes

Attendees:

- Dr. Annette Anderson, Johns Hopkins University
- Terry Ball, Maryland Association of Elementary School Principals
- Tiara Booker-Dwyer, MSDE Office of Leadership Development and School Improvement
- Cheryl Bost, Maryland State Education Association
- Dr. Faith Connolly, Baltimore Education Research Consortium
- Juliann Dibble, Howard County Public Schools
- Dr. Matthew Duque, MSDE Office of Research
- Rick Edwards, Cecil County Public Schools
- Dr. Bonnie Ennis, Wicomico County Public Schools
- Tom Evans, Maryland Association of Secondary School Principals
- Dr. Brian Eyer, MSDE Office of Leadership Development and School Improvement
- Dr. Laurie Henry, Salisbury University
- Lisa Hopkins, Caroline County Public Schools
- Dr. Rhonda Jeter, Bowie State University
- Laura Liccione, MSDE Office of Leadership Development and School Improvement
- Carla McCoy, Baltimore Teachers Union
- Angela Minnici, Mid-Atlantic Comprehensive Center
- Edmund Mitzel, MSDE Office of Leadership Development and School Improvement
- Dr. Gregory Pilewski, Queen Anne's County Public Schools
- Nancy Reynolds, Maryland Association of Boards of Education
- Dr. Kim Rotruck, Frostburg State University
- Jeffery Terziev, Mathematica Policy Research
- Dr. Elias Walsh, Mathematica Policy Research
- Dr. Li Wang, Data Analyst, Office of Leadership Development and School Improvement
- Dr. Mary Young, Prince George's County Public Schools

1) Mid-Atlantic Comprehensive Center reviewed the desired state for evaluation for professional practice that was developed from the September 27, 2018 meeting.

Figure 5: Desired State: Professional Practices for Teachers



2) Workgroup members discussed factors in the desired state for student growth measures.
   a. How should student growth be measured?
      - Student growth has been compliance based, not aligned to professional practice.
      - A better model would be to embed student growth into professional practice. This would be authentic for teachers and provide evidence to improve professional practice.
      - Teachers need a formative assessment piece. Principals and school leaders should be coming in to have conversations about student growth.
      - The teacher needs to see where they are in relation to a student growth number. What is their contribution to that number?
      - This component should have a predictive nature in order to help teachers move in the same direction.
      - What is this teacher doing for the students in their classroom? Good principals hold teachers accountable for the growth of all their students.
      - Principals must be held accountable for the growth of all the students in their building.
      - Student growth must be measured from start of the school year to May.
      - Student growth must take a multi-dimensional approach.
      - Diagnostic tools are needed to drive daily instruction.
      - How will this process be validated over time and is it reliable?

b. What are the intended results?
    - Minimum of 1 year's growth for all subjects
        1. Setting an arbitrary mark is not the answer because it can stifle growth.
        2. Some subjects have better growth measures i.e. English, math. Other subjects are variable i.e. art, music, physical education.
        3. Social emotional learning influences student growth.
        4. Minimum of 1 year of growth should be expected but should not be the ceiling for that child.
    - Use a Plan, Do, Reevaluate model to reflect on practices and make shifts as needed.
    - Growth needs to be realistic and differentiated for student groups. Use of value-added models or student growth percentiles may account for student variation.
    - It is important to keep issues of equity in the forefront when determining student growth outcomes. Is it equitable for all students to achieve the same goal at the same time?

c. What are the non-negotiables?
    - Not a standardized test.
    - Use current school data, not lagging data.
    - Locally flexible – not state directed.
    - Every teacher must be able to see individual impact on student growth.
    - Not a single measure.
    - Easy to understand.
    - Must be used for professional growth not for punishment/compliance.
    - Ensuring that there is an alignment with school improvement, district plan and bridge to excellence.
    - Diagnostic tool to guide instruction.

d. Final Thoughts/Considerations:
    - There should be an alignment to the star rating system including student growth and achievement scores.
    - Change the law to reflect revisions to student growth model.
    - Establish a vehicle to share best practices from around the country and local school systems.
    - Instructionally focused and authentic.
    - Local context matters.
    - Build complete understanding at all levels of the intention of the student growth models (to improve not remove).
    - What happens when the teacher performance does not improve? What is the process?

3) Members from Mathematica Policy Research explained how to compare different student growth measures

    a. Key characteristics of student growth measures were defined.

- Coverage – How many teachers can we use this growth measure for?
- Effort – What is required of school personnel to create this measure?
- Fairness – Does the measure correlate to student characteristics currently available?
- Simplicity – Is the score easily understood?
- Validity – Does it actually measure teacher effectiveness?
- Reliability – Is the result consistently achieved each time?
- Timeliness – Is the data current or lagging?
- Standardization – Is the model implemented with fidelity across all school systems?

b. Review of Student Growth Percentile – Compare all students who achieved the same baseline score.
c. Educator Impact Model - Uses multiple factors to predict the score for the year related to teacher impact.
d. Comparing Student Growth Measures using the key characteristics (see figure 6).

Figure 6: Comparison of Student Growth Models based on the Key Measure Characteristics

	Student Learning Objectives	Student Growth Percentiles	Educator Impact	School-wide
Coverage	High	Low	Low	High
Effort	High	Low	Low	Low
Fairness	Low	Med	High	Low
Simplicity	Low	Med	Low	Med
Validity	Low	Med	High	Low
Reliability	?	Med	Med	High
Timeliness	High	Low	Low	Low
Standardization	Low	High	High	High

4) Workgroup members also discussed school-wide measures and district assessments. Workgroup members were not in favor of school-wide measures because it held teachers accountable for students that they did not teach. Workgroup members were also not in favor of district assessments. This is due to the variability in assessments and the time and resources that it would take to create an assessment for each content area. Workgroup members also felt that this approach would not be fair to teachers of multiple content areas such as elementary school teachers.

5) The workgroup members reviewed five options for teacher evaluations. For each option, notable features were reviewed. Workgroup members were asked to analyze characteristics of each model. Refer to the next page for a summary of models reviewed.

Figure 7: Option 1 - Danielson Framework and Student Growth Percentiles

	Notable Features	Characteristics and Factors to Consider:
Professional Practice	**Relies 100% on *Danielson Framework for Teaching: Six Clusters Supporting High Level Learning;*** ○ Provides a description of planning and instructional skills in promoting high levels of student performance. ○ Content-agnostic and covers 100% of PreK-12 teachers.	❑ Coverage ❑ Effort ❑ Fairness ❑ Simplicity ❑ Validity ❑ Reliability ❑ Timeliness ❑ Standardization
Student Growth	**Tested and Non-Tested Subject Areas:** **Teachers in tested areas**: ○ Student growth percentiles attached to evaluation- **Coverage ~ 15-20% of teachers**; - Uses state assessment data for English language arts and mathematics to find student growth percentiles for test grades only (4^{th}-8^{th} grade). - Lagging data. - Standardized test scores for grades 4-8 only; not high school. **Teachers of Non-Tested areas**: ○ Student learning objectives scores attached to evaluation- Coverage all teachers.	

Figure 8: Option 2 - Danielson Framework and District-Designed Assessments

	Notable Features	Characteristics and Factors to Consider:
Professional Practice	**Relies 100 % on *Danielson Framework for Teaching: Six Clusters Supporting High Level Learning;*** ○ Provides a description of planning and instructional skills in promoting high levels of student performance. ○ Content-agnostic and covers 100 % of PreK-12 teachers.	❑ Coverage ❑ Effort ❑ Fairness ❑ Simplicity ❑ Validity ❑ Reliability ❑ Timeliness ❑ Standardization
Student Growth	**Cluster 5: Successful Learning by All Students** Premise: Emphasizes the improving and enhancing of teacher's professional practices. ○ Adoption of summative assessments or district-designed aligned to standards and goals to measure learning- no lagging data. ○ Allows for continuous and ongoing monitoring and adjustment of student learning, feedback to and from students. ○ Assessment becomes integrated into instruction. ○ Provides a clear pathway on how students are progressing toward learning goals. ○ Content- agnostic; 100 % teachers covered.	

Figure 9: Option 3 - Danielson Framework and Educator Impact/Value Added Model

	Notable Features	Characteristics and Factors to Consider:
Professional Practice	**Relies 100 % on *Danielson Framework for Teaching: Six Clusters Supporting High Level Learning;*** ○ Provides a description of planning and instructional skills in promoting high levels of student performance. ○ Content-agnostic and covers 100 % of PreK-12 teachers.	❑ Coverage ❑ Effort ❑ Fairness ❑ Simplicity ❑ Validity ❑ Reliability ❑ Timeliness ❑ Standardization
Student Growth	**Educator Impact/Value-Added Model** Premise: Uses a statistical model to account for certain student characteristics outside of teacher control. ○ Tested area teachers- use of state assessment data- lagging data. ○ Characteristics could include: - student attendance; - English learner status; - Students with disability status; and - Student mobility.	

Figure 10: Option 4 - Danielson Framework and Student Learning Objectives

	Notable Features- This option was developed during the Workgroup meeting.	Characteristics and Factors to Consider:
Professional Practice	Adopt Charlotte Danielson's Framework for Teaching: Six Clusters with the following recommendations: ○ Districts can adopt a Cluster on a rotating basis each year (i.e., over a three-year evaluation cycle, collaboration between teacher and principal to focus on one Cluster; rotate to a different Cluster the following year)	❑ Coverage ❑ Effort ❑ Fairness ❑ Simplicity ❑ Validity ❑ Reliability ❑ Timeliness ❑ Standardization
Student Growth	Recommendations to: ○ Improve Student Learning Objectives- 100% coverage for teachers; ○ Support: resource, technical assistance, and professional learning from district and/or State; ○ Identify the measurement of fidelity of implementation of the Student Learning Objective	

Figure 11: Option 5 – Workgroup Developed Model

	Notable Features- **This option was developed during the Workgroup meeting.**	**Characteristics and Factors to Consider:**
Professional Practice	Adopt Charlotte Danielson's Framework for Teaching: Six Clusters with the following recommendations: ○ Districts can adopt a Cluster on a rotating basis each year (i.e., over a three-year evaluation cycle, collaboration between teacher and principal to focus on one Cluster; rotate to a different Cluster the following year) ○ Consider adding the Tripod Survey	❑ Coverage ❑ Effort ❑ Fairness ❑ Simplicity ❑ Validity ❑ Reliability ❑ Timeliness ❑ Standardization
Student Growth	Recommendations to: ○ Improve Student Learning Objectives- 100% coverage for teachers; ○ Support: resource, technical assistance, and professional learning from district and/or State; ○ Identify the measurement of fidelity of implementation of the Student Learning Objective	

Workgroup members were instructed to take the options back to their organizations and collect feedback and suggestions to share at the next meeting.

February 15, 2019 Meeting Minutes

Attendees:

- Terry Ball, Maryland Association of Elementary School Principals
- Tiara Booker-Dwyer, MSDE Office of Leadership Development and School Improvement
- Bridgette Blue Laney, Prince George's County Public Schools
- Cheryl Bost, Maryland State Education Association
- Dr. Faith Connolly, Baltimore Education Research Consortium
- Juliann Dibble, Howard County Public Schools
- Dr. Matthew Duque, MSDE Office of Research
- Dr. Bonnie Ennis, Wicomico County Public Schools
- Dr. Brian Eyer, MSDE, Office of Leadership Development and School Improvement
- Tom Evans, Maryland Association of Secondary School Principals
- Linda Gill, Public School Superintendents' Association of Maryland
- Laura Liccione, MSDE Office of Leadership Development and School Improvement
- Carla McCoy, Baltimore Teachers Union
- Angela Minnici, Mid-Atlantic Comprehensive Center
- Edmund Mitzel, MSDE Office of Leadership Development and School Improvement
- Dr. Christopher Morphew, Johns Hopkins University
- Dr. Kristi Murphy, Prince George's County Public Schools
- Nancy Reynolds, Maryland Association of Boards of Education
- Michael Sedgewick, Maryland PTA
- Dr. Carolyn Teigland, Public School Superintendents' Association of Maryland

- Jeffery Terziev, Mathematica Policy Research
- Dr. Elias Walsh, Mathematica Policy Research
- Dr. Li Wang, MSDE Office of Leadership Development and School Improvement
- Dr. Starlin Weaver, Salisbury University

1) Sharing Feedback from Workgroup Members

   Workgroup members were asked to share feedback from their organization feedback related to student growth measures presented during the prior meeting. Many participants shared that student learning objectives are not a successful measure of student growth because of the wide range of variability in school systems across the state. Additionally, SLOs require a lot of time and resources to implement them with fidelity. Participants stated their preference is to maintain local control over their evaluation model, especially if they were experiencing any level of success. A common theme arose surrounding the concern that a change in the student growth model would require further training and resources, which would create a hardship among many local school systems. Local school systems would like more time to implement SLOs and to have more training on the development and implementation of SLOs.

2) Desired State for Student Growth Measures

   The discussion for the desired state of student growth measures emphasized the importance of recognizing the teacher impact on all students with a focus on disadvantaged students in the lowest performing schools. Members of the work group advocated for the need to use multiple measures in the student growth component. Multiple measures will allow for triangulation of data to provide a better picture of teacher performance.

3) Purpose of Teacher Evaluations

   Workgroup members returned to a discussion of the purposes for teacher evaluation. Members selected different purpose for evaluations and the Mid-Atlantic Comprehensive Center matched the purpose to a growth measure. Figure 12 matches evaluation purpose with a growth measure.

Figure 12: Evaluation Purpose and Corresponding Growth Measure(s)

Purpose	Growth model	Student Learning Objectives	Classroom Observation	Teacher Self Reports	Student Surveys
1. Find out whether grade-level or instructional teams are meeting specific achievement goals	X	X			
2. Determine whether a teacher's students are meeting achievement gains	X	X			
3. Gather information to provide teachers with guidance related to identified strengths and areas for growth			X		
4. Examine the effectiveness of teachers in lower elementary grades for which no test scores from previous years are available to predict student achievement			X		
5. Determine supports for new teachers			X	X	X
6. Determine whether a new teacher is meeting performance expectations in the classroom			X		X
7. Gather information to determine what professional learning opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	X	X	X	X	
8. Determine how students perceive a teacher's instructional efforts.					X
9. Determine who would qualify to become a mentor, coach or teacher leader	X	X	X		
10. Gather information on a teacher's ability to work collaboratively with colleagues to evaluate needs of and determine appropriate instruction for at-risk or struggling students			X	X	

Source: Adapted from *A Practical Guide to Designing Teacher Evaluation Systems* - https://gtlcenter.org/sites/default/files/docs/practicalGuideEvalSystems.pdf

4) Review of Measures for Student Growth

A discussion was facilitated by members of Mathematica Policy Research to provide participants the opportunity to address remaining questions about student growth measures. The team presented excerpts from the Student Growth Measures Fact Sheet.

5) Shown below are the results for the consensus activity for 5 options for student growth (see figure 13).

Figure 13: Preliminary Recommendations from Workgroup Members for Student Growth Measures

Option	Professional Practice and Student Growth Measure	Workgroup Recommendations and Comments
1	• Danielson Framework • Student Growth Percentile (tested areas) • Student Learning Objective for non-tested subjects	0 recommendations for this model
2	• Danielson Framework • District Made Assessments for all teachers. • Student growth- Student Growth Percentile or educator impact	0 recommendations for this model
3	• Danielson Framework • Educator Impact mode (tested areas) and • SLO for non-tested subjects	0 recommendations for this model
4	• Danielson Framework • SLO for all teachers	5 recommendations with written comments: ○ Better implementation of SLOs ○ There is no real consensus among the workgroup ○ Pause on SLOs ○ Pause until Maryland Comprehensive Assessment Program is in place ○ Option 1- assessment is changing- SGP- with two different assessments would be suspected and doesn't apply to all teachers ○ Option 2- cost prohibitive for districts ○ Option 3- too complex for a district to implement ○ Option 4- we have already shown that it [SLOs for all] does not work
5	-Open to suggestions and recommendations from Workgroup	7 recommendations with written comments: ○ Option 4.1= growth + impact ○ With changes in SLO and/or process ○ Maintain until clarity is provided [from State] ○ Must be reliable and valid assessment ○ Pause, at this time, with student achievement until new state assessment is fully developed and available for districts to react ○ Develop Danielson and SLO with fidelity ○ Maryland State Education Association: corrections and streamlining of SLOs need to occur in many school systems ○ Maintain local models with great flexibility ○ [SLOs] provide same measure for all subjects/levels ○ Keep current language on professional practice domains ○ Work to remove [student] growth in law
Abstain	3 workgroup members abstained	9 members representing MSDE, Mid-Atlantic Comprehensive Center at West Ed, Mathematica Policy Research abstained

Appendix 2: Summary of Student Surveys for Teacher Evaluation

The table below provides an overview of commonly used student surveys for teacher evaluation. The Maryland State Department of Education does not endorse any particular survey. The list of surveys is supplied for informational purposes only.

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
1. Tripod survey by Ronald Ferguson http://tripodproject.org/	7 Cs of teaching practices • Caring • Captivating • Conferring • Controlling • Clarifying • Challenging • Consolidating	• Using data from a teacher teaching multiple classes, student perceptions are consistent across classes (correlations between .58 and .68). • Additionally, perceptions in one class predict to achievement gains in another class. • Each dimension is highly reliable (.80 and above) and consistent across the school year (.70–.85). • Control and challenge dimensions have the highest correlations with value-added measures (Bill & Melinda Gates Foundation, 2010).	Tailored surveys for Grades K–2, Grades 3–5, and Grades 6–12	36 items[3] in Grades 3–5; 35 items in Grades 6–12	Provides schools, districts, and states with training and support for implementation and analysis. Contact the provider for additional information about cost and implementation.

[1] More rigorous studies should be conducted with all measures reviewed. Many of the validity and reliability studies were not conducted for variations of the tool or across all age ranges for which the developers state the tool is used. Future research should look at how the constructs work across all grades in which the survey is used.

[2] Numbers of items vary depending on the age level. Not all instruments were clear about variations among age levels.

[3] Number of items was not located for the K–2 version.

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
2. My Student Survey by Ryan Balch www.mystudentsurvey.com	• Presenter • Manager • Counselor • Coach • Motivational speaker • Content expert	• One major research study conducted by the survey developer included more than 15,000 students and 900 teachers in Georgia, as part of the state's Race to the Top initiative (Balch, 2012). • The study reported that the constructs were reliable and valid measures. • Student perceptions of teacher behavior predicted to student engagement and academic efficacy, as well as value-added models.	Elementary, middle, and high school versions	63 items	Provides schools, districts, and states with training and support for implementation and analysis. Contact the provider for additional information about cost and implementation.
3. iKnow My Class Survey by Russell Quaglia www.iKnowMyClass.com	• Engagement • Relevance • Relationships • Class efficacy • Cooperative learning environment • Critical thinking • Positive pedagogy • Discipline problems	• A technical report provides information about studies of the tool's validity and reliability, conducted by developers (Bundick, 2011). • The survey was validated with more than 5,000 middle and high school students in the United States and United Kingdom.	Grades 3–5; Grades 6–12	27 items for Grades 3–5 Two forms for Grades 6–12 (20-item form; 50-item form)	Provides schools, districts, and states with training and support for implementation and analysis. Contact the provider for additional information about cost and implementation.

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
4. Questionnaire on Teacher Interaction (QTI) by Wubbels & Levy	• Leadership • Helping/ friendly • Understanding • Student freedom • Uncertain • Dissatisfied • Admonishing • Strict	• The QTI is predicted to student achievement and positive student attitudes (den Brok, Brekelmans, & Wubbels, 2004; Koul & Fisher, 2005; Kyriakides, 2005). • The QTI has been validated in multiple international research studies. The number of items per dimension fluctuated depending on the study (e.g., Goh & Fraser, 1996; Kokkinos, Charalambous, & Davazoglou, 2009; Kyriakides, 2005; Wubbels & Levy, 1991). • Reliability estimates varied depending on the study (Kokkinos et al., 2009; Kyriakides, 2005). • The study found gender differences (girls perceived more cooperative behavior; boys perceived teachers as more oppositional).	Grades 5–6; Grades 6–12	Number varies depending on study	Provides schools, districts, and states with training and support for implementation and analysis. Contact the provider for additional information about cost and implementation.

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
5. 5 Essentials by University of Chicago Consortium on Chicago School Research (CCSR) https://illinois.5-essentials.org/2012/	Two of the five essential elements contain student perceptions of classrooms: • Supportive environment (academic personalism, academic press, peer support for academic work) • Ambitious instruction (math instruction, English instruction, course clarity)	• In a major study on school reform in the 1990s, CCSR validated the majority of measures in the 5Essentials in more than 200 schools in Chicago Public Schools across multiple years (Bryk, Sebring, Allensworth, Luuescu, & Easton, 2010). • CCSR provides customized reports for schools in Chicago and Detroit based on the results of the 5Essentials.	Grades 3–12	33 items related to the subdimensions	Provides schools, districts, and states with training and support for implementation and analysis. Contact the provider for additional information about cost and implementation.
6. SurveyWorks by Rhode Island Department of Elementary and Secondary Education http://www.ride.ri.gov/informationAccountability/RIEducationData/SurveyWorks.aspx	• Teacher practice • Student engagement	• No publicly available or peer reviewed research is documented with the survey.	Grades 4–5; Grades 6–8; Grades 9–12	No information available	No additional information available

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
7. Child Development Project survey by Developmental Studies Center	• Student autonomy and influence • Classroom supportiveness (peers) • Enjoyment of class • Trust in and respect for teachers	• Measures were used within a program evaluation of the social-emotional learning program, Child Development Project (Developmental Studies Center, 2005). • Multiple studies have been conducted to validate the survey. The measures have been used in evaluations with more than 3,000 students (e.g., Solomon, Battistich, Kim, & Watson, 1997) • Reliability of the measures is good (.84–.91), except for enjoyment of class (.66).	Grades 3–5; Grades 6–8	38 items	No additional information available
8. Learner-Centered Battery (LCB) by Barbara McCombs	• Personal/social dimension • Metacognitive/ cognitive dimension • Affective/ motivational dimension • Developmental/ individual differences	• The developer conducted two validation studies with more than 9,000 middle school students (McCombs, Lauer, & Peralez, 1997). • Reliability with measures is good (.71–.91) for the four constructs. • The survey has been used in a variety of research studies (e.g., Meece, 2003).	Grades K–3; Grades 4–5; Grades 6–12	25 items	No additional information available

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
9. Patterns of Adaptive Learning Survey (PALS) by Carol Midgley and colleagues http://www.umich.edu/~pals/manuals.html	• Teacher achievement goals for classroom (mastery, performance–approach, performance–avoidance) • Classroom achievement goals (mastery, performance–approach, performance–avoidance)	• Developers note the validity study of the measures in the manual (Midgley et al., 2000). • Measures are used widely in the study of classroom learning environments and student motivation. • Measures predict to a variety of student outcomes (e.g., efficacy, engagement, regulation). • Measures ask students about teacher behaviors and activities in the classroom that orient students to learning goals.	Elementary, middle, and high school	3–5 items per construct	No additional information available
10. Classroom Life Measure by Johnson & Johnson	• Teacher academic support • Teacher personal support • Peer academic support • Peer personal support • Cooperative learning • Positive goal interdependence • Resource interdependence • Working with heterogeneous peers • Fairness of grading	• Reliability and validity study was originally conducted with 883 students (Johnson, Johnson, & Anderson, 1983). • Reliability estimates ranged by construct from moderate to high (.61–.83). • Select measures have been used in a variety of other research studies (e.g., Patrick, Ryan, & Kaplan, 2007).	Validated in Grades 4–9	39 items	No additional information available

Instrument/ Developer	Constructs Assessed	Validity and Reliability Studies[1]	Grade(s) Used	Number of Items[2]	Additional Information
11. Constructivist Learning Environment Survey (CLES) by Johnson & McClure http://surveylearning.moodle.com/cles/	• Personal relevance • Uncertainty • Critical voice • Shared control • Student negotiations	• Reliability and validity studies have been conducted with more than 1,000 students in science classrooms (reliability coefficients were high, .74–.85) (Nix, Fraser, & Ledbetter, 2005). • The survey was used to evaluate a science reform initiative in Texas. • It has been used mostly in math and science classrooms and has been validated mostly as a teacher self-report. • There have been recent advances to make it a student report along the same dimensions.	Middle and high school	30 items	No additional information available

Works Consulted

1. Rowland, C., Feygin, A., et al (2018). Improving the Use of Information to Support Teaching and Learning Through Continuous Improvement Cycles. American Institutes for Research.
2. Hiring the Best Teachers (2003). Educational Leadership. Vol. 60(8), pages 48-52.
3. Teachers Matter: Understanding Teachers' Impact on Student Achievement. Santa Monica, CA: RAND Corporation, 2012. https://www.rand.org/pubs/corporate_pubs/CP693z1-2012-09.html.
4. Boren, M., Brookhart, E., Vaishnav, A., (2019). Educator Effectiveness Final Progress Monitoring Report: Maryland Fall/Winter 2018-19. Southern Regional Education Board and Education First.
5. C. Kirabo Jackson, 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes," Journal of Political Economy, vol 126(5), pages 2072-2107.
6. Center on Great Teachers and Leaders (2014). Practical Guide to Designing Comprehensive Teacher Evaluation Systems: A Tool to Assist in the Development of Teacher Evaluation Systems.
7. Data Quality Campaign (January 2019). Growth Data: It Matters, and It's Complicated.
8. Refining Student Growth Measures in Maryland's Educator Evaluation System. Presented October 25, 2018, REL Mid-Atlantic for Mathematica Policy Research.
9. Danielson, C. (2015). Framing Discussion About Teaching. Educational Leadership, 2015. v. 72(7) 28-41.
10. New America. (March 2016). Beyond Ratings: Re-Envisioning State Teacher Evaluation Systems as Tools for Professional Growth.
11. Maryland Report Card (new release December 2018). http://reportcard.msde.maryland.gov/HelpGuides/Maryland_ReportCard_User_Guide.pdf
12. The Danielson Group (2015). The Framework for Teaching Clusters: Six Clusters Supporting High Level Learning.
13. Center on Standards and Assessment Implementation at WestEd (n.d.). Assessment Design Toolkit: https://www.csai-online.org/node/1997.
14. The Center for Education Policy Research at Harvard University (2011). Value-added measures: How and why the strategic data project uses them to study teacher effectiveness.
15. Regional Educational Laboratory at Mathematica Policy Research. Fact Sheet: Measuring progress in the classroom: How do different student growth measures compare? (November 2018).
16. Maryland State Department of Education (October 2018). Principal Evaluation Guidebook.
17. Center on GREAT TEACHERS & LEADERS at American Institutes for Research, & CCSSO. The Professional Standards for Educational Leaders (PSEL) 2015 and the Interstate Leaders Licensure Consortium (ISLLC) Standards 2008: A Crosswalk (2016).