

**Report of the
Regents Task Force on Teacher and Principal Effectiveness**

April 4, 2011

Table of Contents

OPENING STATEMENT OF THE TASK FORCE	5
EXECUTIVE SUMMARY	7
BACKGROUND ON THE REGENTS TASK FORCE	7
SUMMARY OF KEY PROVISIONS OF EDUCATION LAW §3012-C	7
IMPLEMENTATION TIMELINE	8
OVERVIEW OF THE REPORT	9
<i>Student Growth Measures: Teachers</i>	9
<i>Locally-Selected Measures of Student Achievement: Teachers</i>	12
<i>Student Growth Measures: Principals</i>	13
<i>Locally-Selected Measures of Student Achievement: Principals</i>	14
<i>Other Measures of Effectiveness: Teachers (60%)</i>	15
<i>Other Measures of Effectiveness: Principals (60%)</i>	16
<i>Composite Scoring Bands</i>	18
<i>Subcomponent Scoring</i>	19
MEASURING STUDENT GROWTH IN NON-TESTED GRADES AND SUBJECTS	20
OVERVIEW OF OPTIONS AND CRITERIA	20
1. DEVELOP STATEWIDE GROWTH MEASURES USING EXISTING STATE TESTS	23
<i>1a – Use Existing State Assessments in Aligned Content Areas</i>	23
<i>1b – Use Existing State Assessments in Non-Aligned Content Areas</i>	23
<i>1c – Developing New NYS Assessments for All Currently Non-Tested Content and Subject Areas</i>	24
2. ALLOW DISTRICTS TO CHOOSE FROM A MENU OF OPTIONS FOR ASSESSING STUDENT GROWTH.....	25
<i>2a - District-Developed Assessments</i>	25
<i>2b - Regionally-Developed Assessments (through BOCES or other Regional Consortia)</i>	26
<i>2c - Commercially Available Assessments Identified and Selected By the District</i>	27
<i>2d - Commercially Available Assessments Selected by the District from a Menu of Assessments Identified as Rigorous by the State</i>	28
<i>2e - Group Metric from State Assessments</i>	29
<i>2f - District-Wide Growth Goals for Selected Assessment</i>	30
<i>2g - Performance Assessments Accepted By Professions</i>	30
STUDENT GROWTH MEASURES FOR PRINCIPALS	31
<i>Near-Term Measures in Addition to Results of Growth/Value-Added Model</i>	31
<i>Growth Within Student Subgroups</i>	32
<i>Student Growth in Non-Tested Subjects</i>	32
LOCALLY-SELECTED MEASURES OF STUDENT ACHIEVEMENT	34
RECOMMENDED.....	35
FURTHER DISCUSSION REQUIRED.....	37
NOT RECOMMENDED.....	38
TASK FORCE DISCUSSIONS	39
<i>Issue 1: Local comparability</i>	39
<i>Issue 2: Rigor</i>	40
<i>Issue 3: Individually Teacher-Created Assessments</i>	41
<i>Issue 4: Scope of Collective Bargaining</i>	44
<i>Task Force positions</i>	44
LIST OF APPENDICES.....	44
OTHER MEASURES OF EFFECTIVENESS: TEACHERS (60%)	45
THE TEACHING STANDARDS.....	46
PROPOSED CRITERIA FOR TEACHER PRACTICES/ RUBRICS	46
ASSESSMENT APPROACHES FOR TEACHERS.....	47
<i>Classroom Observations</i>	47
<i>The Use of Rubrics</i>	48

OTHER MEASURES OF EFFECTIVENESS: PRINCIPALS (60%)	50
INTRODUCTION	50
OVERVIEW OF RECOMMENDATIONS	51
STANDARDS FOR PRINCIPAL EVALUATIONS	53
<i>Adopt ISLLC Standards</i>	<i>53</i>
<i>Allow Districts to Apply for a Variance</i>	<i>54</i>
<i>Develop State Standards</i>	<i>54</i>
RUBRICS TO ASSESS PROFICIENCY	55
<i>Provide a Menu of State-Approved Choices and a District Variance Process</i>	<i>55</i>
<i>Allow for the Free Choice of Any Rubric</i>	<i>56</i>
<i>Identify a Single Statewide Rubric</i>	<i>56</i>
REQUIRED ELEMENTS OF A PRINCIPAL PERFORMANCE EVALUATION.....	56
<i>Feedback from the Supervisor Plus Two Additional Stakeholders</i>	<i>57</i>
<i>Standardized Rubric.....</i>	<i>57</i>
REQUIRED MEASURES OF PRINCIPAL PERFORMANCE	57
<i>Broad Assessment of Principal Leadership and Actions.....</i>	<i>58</i>
<i>Assessment of Principal’s Skill in Supervising and Developing Teachers.....</i>	<i>58</i>
<i>Progress Against Individual Professional Growth Goals.....</i>	<i>59</i>
<i>Progress Against Ambitious and Measurable School Academic or Learning Environment Goals.....</i>	<i>61</i>
SOURCES OF EVIDENCE	61
<i>School Visits.....</i>	<i>62</i>
<i>Formalized and Standardized Feedback from Students, Parents, Teachers, Community Stakeholders</i>	<i>62</i>
<i>Principal Self-Reflection.....</i>	<i>62</i>
<i>Review of School Documents, Systems, and Records.....</i>	<i>63</i>
<i>Results and Findings from Existing State Accountability Processes.....</i>	<i>63</i>
<i>Other Sources as Determined Locally</i>	<i>63</i>
WEIGHTING OF EVIDENCE	63
“METRICS”: TEACHER OF RECORD, SCORING, AND VALUE-ADDED	64
CONTEXT.....	64
TEACHER OF RECORD POLICY	64
SUBCOMPONENT SCORING (STATE GROWTH, LOCAL ASSESSMENT, AND OTHER 60%).....	65
COMPOSITE SCORING BANDS	68
GROWTH AND VALUE-ADDED MODELING	70
<i>Student Growth Model</i>	<i>71</i>
<i>Value-Added Model.....</i>	<i>71</i>
<i>Teacher and Principal Student Growth (or VA) Scores.....</i>	<i>71</i>
<i>State Data Elements.....</i>	<i>72</i>
<i>Status Report on Growth/VA modeling</i>	<i>74</i>
EVALUATOR TRAINING.....	75
<i>Content of Evaluator Training.....</i>	<i>75</i>
<i>Instructional Delivery</i>	<i>76</i>
<i>Timeline.....</i>	<i>76</i>
<i>Regulations and Guidance</i>	<i>77</i>
PROFESSIONAL DEVELOPMENT FOR TEACHERS AND PRINCIPALS.....	78
APPENDICES.....	80
APPENDIX A—LIST OF TASK FORCE MEMBERS AND WORKGROUP ASSIGNMENTS	80
APPENDIX B—ADVISORS, PRESENTERS, STAFF, AND FELLOWS	83
APPENDIX C—WORK CITED AND WORKS REVIEWED.....	84
APPENDIX D—CHAPTER 103 OF THE LAWS OF 2010	88
APPENDIX E—SYNTHESIS DOCUMENT OF THE LOCALLY-SELECTED MEASURES WORKGROUP FROM JANUARY 2011 MEETING	96
APPENDIX F—WORKGROUP OPINION ON STATE TESTS FOR 20% LOCAL	99

APPENDIX G—SCENARIOS FOR USE OF STATE TEST RESULTS IN THE OTHER 80%, BY MEMBER OF METRICS WORKGROUP, MARCH 24, 2011	101
APPENDIX H—TEACHER-CREATED ASSESSMENTS, BY MEMBER OF WORKGROUP ON LOCALLY-SELECTED MEASURES, JAN 30, 2011	103
APPENDIX I—LOCAL BARGAINING ISSUES, BY ONE MEMBER OF WORKGROUP ON LOCALLY-SELECTED MEASURES	106
APPENDIX J—PROPOSED FRAMEWORK FOR PRINCIPAL EVALUATION.....	110
APPENDIX K—60% PRINCIPALS SUMMARY TABLE OF PROPOSED ELEMENTS AND MEASURES	111

Opening Statement of the Task Force

We, the members of the Regents Task Force on Teacher and Principal Effectiveness (Task Force) are pleased to present our work to the Board of Regents. As representatives of our stakeholder groups, our one unifying purpose is – to develop a comprehensive teacher and principal evaluation system that will improve teaching practice to advance learning for all students. The process has been challenging at times. Our discussions have been spirited and disagreements have produced new ideas, insights and determinations. Our commitment has resulted in a true collaboration.

This new system will be a comprehensive restructuring of how teachers and principals are evaluated and New York State is leading the way. It is all new, and there is no existing, comparable system that can provide a blue-print for us to follow. What we do know, and all stakeholders share, is the understanding that the new system must be fair, transparent and result in meaningful evaluations for teachers and principals. It must be comprehensible to those being evaluated and also to the public. While there is an ideal and a vision to our work, there is also a practical aspect in its implementation. For meaningful reform to occur, it must be flexible to ensure it is embraced at the local level. We know that no two districts are alike in population, geography, size, capacity, or economics. The collaborative approach used by the Task Force to reach consensus is a model for how the new evaluation system should be implemented in our schools. With clear and ambitious definitions, rigorous rubrics aligned to New York Teaching Standards or grounded in the Interstate School Leaders Licensure Consortium (ISLLC) standards for principals, and the integrity of professional educators, we can implement a meaningful evaluation system.

We are pleased to report that we have reached consensus on many of the issues we were charged to address. However, there is more that must be done. The outstanding issues will be the most difficult ones to address, and every member of the Task Force is committed to getting it right. The results of these deliberations and the policy decisions the Regents must make have life-changing consequences for all educators involved in school districts across the state. All the elements of the evaluation system are interconnected, and each must be fully operational if the system is to work properly.

The Task Force is highly concerned with the ability of districts to implement the system with fidelity in the first year, and urges the Board of Regents to consider that as it creates policy. Under even the best of circumstances, the implementation of new policies can be challenging, and this year we do not have the best of circumstances. The Regents must take into account the cost implications to districts as it considers this new policy. Districts will need time to build capacity to sustain these changes. We are committed to

this reform, but we are concerned that too many issues are being treated as if they can be solved along the way. This approach will result in frustration, loss of faith and set the system up for failure. If the implementation of the new system is compromised, we may miss our best opportunity to create a meaningful evaluation system that will positively impact student learning.

The Task Force is committed to the successful implementation of the new evaluation system. Our initial report and recommendations are now in your hands, and we stand ready to continue our work on the outstanding issues as the plan continues to evolve.

We would also like to acknowledge and thank the staff of the New York State Education Department (NYSED). We are grateful for their work and commitment to the successful roll-out of this new system.

We thank the Board of Regents for this opportunity to engage all stakeholders to improve teaching and learning for students across New York State.

March 2011

Executive Summary

Background on the Regents Task Force

The Task Force is an advisory committee convened pursuant to Chapter 103 of the Laws of 2010 to provide recommendations to the Commissioner of Education and the Board of Regents on the development of regulations to implement Education Law §3012-c, New York State’s teacher and principal evaluation statute. The Task Force—made up of representatives of teachers, principals, superintendents of schools, school boards, school districts and boards of cooperative educational services (BOCES) officials, and other interested parties—has been meeting regularly since September 2010. Board of Regents Vice Chancellor Milton Cofield and Regent Lester Young served as co-chairs of the Task Force. (See Appendix A—List of Task Force Members and Workgroup Assignments.)

Throughout its deliberations, the Task Force has been supported by the active participation of teams of research advisors, and numerous experts have made presentations to the Task Force. Research and best practice examples were disseminated and discussed at length. (See Appendix B—Advisors, Presenters, Staff, and Fellows and Appendix C—.)

Task Force members were divided into workgroups as needed to consider particular aspects of Education Law §3012-c and to provide input and guidance on certain issues. (Appendix A—List of Task Force Members and Workgroup Assignments.) Each workgroup was supported and facilitated by NYSED staff and Regents Research Fund fellows. Each chapter in this report was originally based on the work of the workgroups and further developed with input from the full Task Force.

Summary of Key Provisions of Education Law §3012-c

Under Education Law §3012-c (see Appendix D—Chapter 103), each teacher and principal will receive an annual professional performance review (APPR) resulting in a single composite effectiveness score and a rating of “highly effective,” “effective,” “developing,” or “ineffective.” The composite score will be determined as follows:

- 20% - student growth on state assessments or a comparable measure of student growth (increased to 25% upon implementation of a value-added (VA) growth model)

- 20% - locally-selected measures of student achievement that are determined to be rigorous and comparable across classrooms (decreased to 15% upon implementation of value-added growth model)
- 60% - other measures of teacher/principal effectiveness

By law, the APPR is required to be a significant factor in employment decisions such as promotion, retention, tenure determination, termination, and supplemental compensation, as well as a significant factor in teacher and principal professional development. If a teacher or principal is rated “developing” or “ineffective,” the school district or BOCES is required to develop and implement a teacher or principal improvement plan (TIP or PIP). Tenured teachers and principals with a pattern of ineffective teaching or performance – defined by law as two consecutive annual “ineffective” ratings – may be charged with incompetence and considered for termination through an expedited hearing process. The law provides further that all evaluators must be appropriately trained and that appeals procedures are to be locally established.

Implementation Timeline

- **July 2010** – New collective bargaining agreements must be consistent with the requirements of Chapter 103 of the Laws of 2010
- **September 2010** – Task Force begins meeting
- **Spring 2011** – Task Force reports to Board of Regents and Regents discuss and adopt regulations
- **2011-2012 school year** – New performance evaluation system takes effect for classroom teachers of common branch subjects, ELA or math in grades 4-8 along with their respective building principals
- **2012-2013 school year** – New performance evaluation system goes into effect for all teachers and building principals
- **2012-2013 school year and thereafter** – Implementation of teacher and principal improvement plans, as appropriate, and implementation of a Regents-approved value-added growth model to be used for the teacher and principal performance evaluation system

Overview of the Report

At the request of the Board of Regents, the Task Force report examines a wide range of policy options, documenting Task Force members’ views and noting implementation considerations. Many of the policy options were advanced by members of the Task Force themselves, while others are based on the evaluation practices and proposals of other states and school districts; still others were identified by NYSED staff, Regents Research Fund fellows, or expert advisors. While the policy options come from a variety of sources, the views and implementation considerations expressed in this report are solely those of the Task Force members. Because the Task Force members represent diverse constituencies, the report expresses multiple perspectives on many of the policy options. A summary of the options and Task Force positions follows.

Student Growth Measures: Teachers

	ELA/Math 4-8 (2011-12 and beyond)	All Classroom Teachers (2012-13 and beyond)
<p>Growth on State Assessments</p> <p>20 percent (25% with approved VA model)</p>	<p>Result of student growth model (value-added if approved for use beginning 2012-13)</p>	<ul style="list-style-type: none"> • Task Force accepts that if NY State adds to its assessment program in additional grades and subjects and/or develops a growth or value-added model for these or other existing state assessments that are approved by the Regents, then the results could be used for this portion of the affected teachers evaluations provided there is empirical evidence for the use of the model for teacher and principal evaluation. • Task Force accepts that, for the purposes of comparability and rigor, adding assessments in “core subject areas” ELA, Math, Science and Social Studies with same subject-area sequenced tests could be the best solution. This would include: <ul style="list-style-type: none"> ○ 6-8 Science and Social Studies if assessments were added ○ Grades 9 and 10 ELA if they sequence clearly with 3-8 ELA. However, many concerns remain with

		<p>adding assessments.</p> <ul style="list-style-type: none"> • The Task Force had mixed opinions but was generally skeptical that growth/value-added models can be constructed for Regents exams even if they follow same-subject middle school assessments or same-subject Regents. Empirical evidence of the validity of a growth/value-added modeling approach would be required. • The Task Force generally does not believe that different-subject post-tests should be used alone as “pre-tests”; for example, 7th grade math for 8th grade science.
<p>Growth in non-tested grades and subjects, i.e., where there is no state assessment</p>	<p>N/A</p>	<p>Some Task Force members want the State to establish a state-wide method of determining student growth for all grades and subjects.</p> <p>Others feel that, in any remaining grades and subjects, state-wide approaches are not feasible or as desirable as providing districts some flexibility to choose, from the options below, the growth measure that best meets the intended purpose. Some options are more appropriate for some grades/subjects than others, and the Task Force articulates pros and cons for each option in the full report.</p> <ul style="list-style-type: none"> a) District-developed assessments b) Regionally developed assessments (through BOCES or other regional consortia) c) Commercially available assessments identified and selected by the district d) Commercially available assessments selected by the district from menu of assessments identified as rigorous by the state e) Group or team measures based on state

		<p>assessments, but only for instances in which the teachers involved in the group measure actually function as a team</p> <ul style="list-style-type: none">f) District-wide growth goal setting process with a variety of assessmentsg) Task Force does NOT recommend that performance assessments like the NYSSMA music evaluations or the NOCTI CTE assessments be used for teacher or principal evaluation without significant work.
--	--	--

Locally-Selected Measures of Student Achievement: Teachers

Locally-Selected Measures of Student Achievement for Teachers of ELA/Math 4-8 20 percent (15% after VA model) Can be growth or achievement	
<p>General Recommendations</p> <ul style="list-style-type: none"> This component of the evaluation system should be based on assessments that are not standardized State tests. These locally-selected measures must add value to classroom instruction. All options must be verified for comparability, validity, rigor, degree to which the assessment can aligned to State standards, and degree to which the assessment(s) match classroom instruction. 	
Teacher Options	Comments
1. Pre-approved menu of acceptable 3 rd party assessments	The Task Force agrees that this is one viable option for a locally selected assessment. There is also the risk of adding more high-stakes <u>tests</u> (vs. assessments) and potential unintended consequences.
2. Districts issue RFP for assessment and verify comparability and rigor.	The Task Force agrees that this is one viable option for a locally selected assessment. This would require an RFP process and has fiscal implications. These assessments would be standardized across the district. There are the same potential risks as with option 1.
3. Districts develop their own assessments and verify comparability and rigor	The Task Force agrees that this is one viable option for a locally selected assessment. This option is resource-intensive but offers opportunities for professional development and building local capacity. These assessments would be standardized across the district.
4. School-wide, group or team assessments (group metric)	The Task Force thinks that this is one viable option for a locally selected assessment but not likely for ELA and Math 4-8. A district may decide to use a school-wide or grade-wide assessment to emphasize team-based teaching or address a district/school goal.
5. Districts allow schools to utilize teacher-created assessments	The Task Force supports teacher created assessments with a mechanism in place to verify comparability and rigor (a rubric tool). Subjectivity in the system cannot be fully controlled for. This option encourages collaboration with the principal and professional development.

Options that were discussed by the Task Force, but not included in the recommendations are presented on Page 38.

Student Growth Measures: Principals

	Elementary/Middle (with 4-8 in 2011-2012; all in 2012-2013 and beyond)	High School (2012-2013)
<p>Statewide Measures of Growth</p> <p>20 percent (25% after VA model)</p>	<ul style="list-style-type: none"> ○ Result of student-growth/value-added model for all tested grades/subjects; add grades or subjects as value-added model applies 	<ul style="list-style-type: none"> ○ Result of student-growth or value-added model as applied to English and Math Regents performance; add subjects as value-added model applies <p><i>Potential near-term measures that could be explored:</i></p> <ul style="list-style-type: none"> ○ Progress toward graduation rate against predicted metric ○ Credit accumulation ○ Course grades/failure rates

Locally-Selected Measures of Student Achievement: Principals

Locally-Selected Measures of Student Achievement for Principals	
20 percent (15% after VA model)	
Can be growth or achievement	
General Recommendations <ul style="list-style-type: none"> This component of the evaluation system should be based on assessments that are not standardized State tests. These locally selected measures must add value to classroom instruction. All assessments must pass muster for rigor (by rubric in standard) and comparability across schools in a district. 	
General Recommendations	Comments
Based on the local assessments used for teachers/multiple measures.	Does not need to be an aggregate of all grades and subjects, but could have focus on a district priority such as ELLs or writing. There must be a balance between the big picture and focused goals. Some goals may span multiple years.
The scoring would be done locally.	The evaluator of principals (presumably superintendents) provides the final rating to the SED, not the data that the SED then feeds into a formula.
Options	Comments
Locally selected assessment that illustrates the achievement level of ONE identified group of students (ELL, a grade level, Special Education, Free and Reduced Lunch, or other) on ALL or MANY content/subject areas. and/or Locally selected assessment(s) that illustrate the achievement level of ALL students on ONE identified content/subject area.	The assessment(s) could be the same as the ones used for teachers of the identified students or an alternative assessment for this purpose. The assessment(s) could be a single test of achievement, growth from one year to the next, or pre/post improvement.

Options that were discussed by the Task Force, but not included in the recommendations are presented on Page 38.

Other Measures of Effectiveness: Teachers (60%)

Standards	There is consensus that the evaluation system should be grounded in the New York State Teaching Standards.
Choice of rubrics to assess proficiency against standards	The regulations and guidance on teacher evaluation should reflect flexibility in the choice of rubrics to assess teacher proficiency. There is consensus around the State developing a list of pre-approved rubrics from which a district could choose, as long as there is a variance process allowing districts flexibility to use a pre-existing rubric currently used in the district.
Required elements	<ul style="list-style-type: none"> • Multiple Sources of Feedback: There was consensus that multiple measures may be included (e.g., multiple observers, additional options negotiated locally, consistent with collective bargaining). • Observation by principal or other administrator is required.
District options	<p>In addition to the requirements above, the Task Force reached consensus (except as noted) that the following options can be considered, subject to local bargaining agreements. This list is meant as a starting point and should not be thought of as all-inclusive.</p> <ul style="list-style-type: none"> • Classroom Observation: <ul style="list-style-type: none"> ○ Observation by trained in-school peer reviewers ○ Observation by independent, trained observers (<i>mixed views</i>) • Video-taped classroom sessions (<i>mixed agreement: more support for use for professional development, personal growth/reflection, "norming" evaluators than as a method of observing for evaluations</i>) • Structured review of student work • Teacher artifacts using "portfolio" or "evidence binder" processes • Feedback from students, parents, and/or other teachers using structured survey tools • Individual professional growth plan • Teacher self-reflection
Training	There was consensus that all evaluators should be trained
Professional Development	Individualized professional development plans should include professional growth goals that are individually established by both the teacher and the principal or other appropriate personnel

Other Measures of Effectiveness: Principals (60%)

Standards	There is broad but not complete agreement that the evaluation system should be grounded in the 2008 ISLLC standards
Choice of Rubrics to Assess Proficiency	The regulations and guidance on principal evaluation should reflect flexibility in the choice of rubrics to assess the data on principal proficiency. There is mixed reaction to whether this should be a menu or free choice.
Required Elements of a Principal Performance Evaluation	<p>A principal performance evaluation should:</p> <ul style="list-style-type: none"> • broadly assess principal leadership and management actions; • occur annually; • yield results that allow for the differentiation of principal performance and have the capacity to differentiate between the four HEDI titles; • promote collaboration, ongoing communication, timely feedback, and trust between the supervisor and the principal being evaluated; • be based on multiple measures that are state defined; • include feedback from the supervisor plus two additional stakeholders; • include at least two sources of evidence for each measure, with feedback being one source of evidence; • allow flexibility for districts to take into consideration factors such as a principal's tenure status, years of experience, years serving in a particular school, and local context; and • be achieved using a standardized rubric, in which all evaluation tools and data collection sources are – to the extent practicable – research-based and psychometrically strong, with demonstrable validity and reliability.
Required Measures of Principal Performance	<p>The assessment should include measures of the following:</p> <ul style="list-style-type: none"> • Broad assessment of principal leadership and management actions (through supervisor assessment of principal actions) • Broad assessment of principal leadership and management actions (through feedback rubric) • Assessment of a principal's skill in supervising and developing teachers

	<ul style="list-style-type: none"> • Progress against individual professional growth goals • Progress against ambitious and measurable school academic or learning environment goals
Sources of Evidence	<p>There should be multiple sources of evidence for each measure. The following optional sources are suggested:</p> <ul style="list-style-type: none"> • School visits • Formalized and standardized feedback over and above the two required sources (students, parents, teachers, community stakeholders) • Principal self-reflection • Review of school documents, systems, and records • Results and findings from existing state accountability processes • Other sources as determined locally
Weighting of Evidence	<ul style="list-style-type: none"> • Evidence should be scored holistically, with local districts determining the weight to be given to each element. • The regulations should not mandate specific weights or numbers/ranges of points for particular measures within the 60%; districts should be free to determine how the measures within the 60% are scored.

Composite Scoring Bands

The Task Force considered numerous options for scoring bands at the composite level. After much deliberation, the group was able to narrow options for scoring bands as reflected in the table below, but did not reach consensus. The group discussed extensively the pros and cons to using a 100 point scale and how this can easily be converted into a “grade” for a teacher or principal, but did not come to consensus about whether an alternative was more appropriate. Further details about the extent of this discussion are included in the report. Assuming a 0-100 range,

Rating Category	Band	Note
Ineffective (I)	0-X	X must be at least 39; some supported 49, and could go as high as 65.
Developing (D)	X-Y	X can be no lower than 40. Y is between 64 and 74.
Effective (E)	P-85	P is between 65 and 85.
Highly Effective (H)	86-100	Although some supported lower limit as high as 91.

Subcomponent Scoring

The Task Force did not complete discussion on subcomponent scoring, although they agreed on a number of principles described in this report. They recognized that they are split on how prescriptive the state should be on the locally-selected measures of student achievement and “other 60%” subcomponents. There is also a split between those who believe that it is the responsibility of the state to determine exactly how growth in non-tested subjects will be measured and how the growth score will be calculated, and those who believe districts should have flexibility to choose from a menu of options for measuring growth in non-tested subjects.

The Task Force achieved general consensus that text descriptions (along the lines of the chart below) of four levels of performance in each category are helpful to guide further discussion on scoring approaches.

Level	Growth	Locally-selected assessment results	Other (Teacher and Leader standards)
L1	Educator’s students’ results are well-below state average for similar students.	Educator’s results do not achieve expectations for student learning.	Not approaching standard on most areas associated with district priorities within standards.
L2	Educator’s students’ results are below state average for similar students.	Educator’s results partially achieve expectations for student learning.	Meets some standards. Needs improvement in areas associated with district priorities.
L3	Educator’s students’ results meet state average for similar students.	Educator’s results achieve expectations for student learning.	Meets all or almost all standards associated with district priorities, may exceed some and may be in need of improvement in areas that are lower priority for the district.
L4	Educator’s students’ results are well above state average for similar students.	Educator’s results exceed expectations for student learning.	Meets all standards and exceeds many standards within district priorities.

MEASURING STUDENT GROWTH IN NON-TESTED GRADES AND SUBJECTS

Overview of Options and Criteria

New York’s teacher and principal evaluation statute provides that 20% of the evaluation is to be based on student growth on state assessments, “*or a comparable measure of student growth if such growth data is not available.*” New York currently has statewide post-testing only in grades 4-8 English language arts and math, so during the first year of implementation, the system is designed to apply just to teachers in those subjects/grades and their principals (assuming that a previous score on a state assessment could be used as the “pre-test” for the following year).

Fourteen percent of current NYS teachers are teachers of ELA and Math in grades 4-8 and an additional 7% teachers may be included who are teachers of special education and ELL/bilingual depending on the assessments their students take and the state’s ability to identify multiple teachers of record. That leaves 79% who fall into the category of Non-Tested Grades and Subjects. To measure student growth for this number and variety of teachers is complex. The Task Force has examined two categories of solutions:

1. Develop statewide growth measures using existing state tests. The Task Force considered 3 sets of subjects for which the state can develop student growth measures using existing state tests, as follows:

Subjects	Possible Approach
1a. Tested subjects for which there is a State post-test and one or more State post-tests in the same content area that could be used as pre-tests,	Use existing state assessments in aligned content areas
1b. Tested subjects for which there is a State post-test and one or more State post-tests that could be used as pre-tests in other content areas, and	Use existing state assessments in non-aligned content areas
1c. Currently non-tested subjects for which new state assessments will be designed.	Develop new NYS assessments for all currently non-tested content and subject areas

	Teachers and Principals beyond grades 4-8 ELA/math (Principals also discussed separately at end of chapter)
Growth on State Assessments 20 percent (25% with approved VA model)	<ul style="list-style-type: none"> • Task Force accepts that if NY State adds to its assessment program in additional grades and subjects and/or develops a growth or value-added model for these or other existing state assessments that are approved by the Regents, then the results could be used for this portion of the affected teachers' evaluations provided there is empirical evidence for the use of the model for teacher and principal evaluation. • Task Force accepts that this may be the best solution for the purposes of comparability and rigor and that this could be done in "core subject areas" ELA, Math, Science and Social Studies with same subject-area sequenced tests. This would include: <ul style="list-style-type: none"> ○ 6-8 Science and Social Studies if assessments were added ○ Grades 9 and 10 ELA if they sequence clearly with 3-8 ELA. • The Task Force had mixed opinions but was generally skeptical that growth/value-added models can be constructed for Regents exams even if they follow same-subject middle school assessments or same-subject Regents. Empirical evidence of the validity of a growth/value-added modeling approach would be required. • The Task Force generally does not believe that different-subject post-tests should be used alone as "pre-tests"; for example, 7th grade math for 8th grade science.

2. Allow districts to choose from a menu of options for assessing student growth. Some Task Force members want the State to determine state growth for all grades and subjects. Others feel that, in any grades and subjects where the State does not have a growth or value-added model, districts should be allowed some flexibility to choose a growth measure; the Task Force considered a menu of seven options that would afford flexibility and feasibility to align best assessment choices to the unique and specific needs of the large number of subjects that fit within the non-tested areas. Note that these options are similar to the options available to districts for "local measures of student achievement." A Task Force member suggested the possibility that the points for state growth could be redistributed among the other two subcomponents (locally-selected measures of student achievement and 60% other measures), but this option is not included in the table below because it was not discussed by the full Task Force and because there is some doubt about the legality of any plan where student achievement would count for less than 40%.

Growth in non-tested grades and subjects, i.e., where there is no state assessment	<p>Some Task Force members want the State to determine state growth for all grades and subjects.</p> <p>Others feel that districts should be allowed the flexibility to choose, from the options below, the growth measure that best meets the intended purpose. Some options are more appropriate for some grades/subjects than others, and the Task Force articulates pros and cons for each option in the full report.</p> <ul style="list-style-type: none"> a) District-developed assessments b) Regionally developed assessments (through BOCES or other regional consortia) c) Commercially available assessments identified and selected by the district d) Commercially available assessments selected by the district from menu of assessments identified as rigorous by the state e) Group or team measures based on state assessments, but only for instances in which the teachers involved in the group measure actually function as a team f) District-wide growth goal-setting process with a variety of assessments g) Task Force does NOT recommend that performance assessments like the NYSSMA music evaluations or the NOCTI CTE assessments be used for teacher or principal evaluation without significant work.
---	--

A major assumption of the Task Force is that a district could defend the selections made as options are chosen. A district measure of growth may apply to all teachers if a district option is determined to be the very best choice for a pre-test, post-test, or both pre- and post-test. (Some Task Force members believe that districts should even be allowed to use a measure of their own choosing in lieu of a state assessment.)

The Task Force suggests that, with additional time on task, it could specify recommendations that align specific instructional circumstances with the appropriate assessment choice.

The Task Force evaluated all foregoing options according to the following criteria:

- Consistent with the new Teacher and Principal Evaluation Statute

- Validity
- Reliability
- Comparability
- Implementation
- Cost
- Rigor
- Effect on Instruction

1. Develop Statewide Growth Measures Using Existing State Tests

1a – Use Existing State Assessments in Aligned Content Areas

In subjects for which there is a State post-test and one or more State post-tests that could be used as pre-tests in the same content area, the Task Force accepts the approach of measuring growth using those existing state assessments as possible short-term and long-term options. The State Education Department would determine the point value to be assigned to each teacher for growth.

For this purpose, “same content area” is defined as tests that fall within the same sequence. For example, grade 7 and 8 math tests would be in the same testing sequence. This approach may apply to teachers of Grade 4-8 ELA and Grade 4-8 Math.

The Task Force considered this approach to be consistent with NYS law, that it would supply valid and reliable content-specific data aligned with NYS Standards, and that NYS assessments provide multiple years of existing data with predictive power.

The Task Force noted, however, that the use of this approach could only be applied where sufficient evidence exists that the growth model for a given grade or subject is empirically rigorous.

1b – Use Existing State Assessments in Non-Aligned Content Areas

In subjects for which there is a State post-test and one or more State post-tests in other, non-aligned, content areas that could be used as pre-tests, the majority of Task Force members, at this time, do not endorse the approach of using the previous state assessment to determine student growth.

“Different content area” is defined in one of two ways: 1) tests in different testing programs (e.g. using Grade 8 ELA as pre-test to measure growth in Regents Global

History) or 2) tests in separate sequences of a broader subject area (e.g. Regents Earth Science as pre-test to measure growth in Regents Physics).

Assuming the growth model were sufficiently rigorous and that the comparability and correlations between pre- and post-tests in non-aligned subject areas were validated by research, some members of the Task Force considered this approach to be consistent with NYS law and that it would supply valid and reliable content-specific data aligned with NYS Standards.

However, given the paucity of available research on such an approach, the sense of many members of the Task Force is that practitioners may not view the approach as credible. The Task Force also noted that implementation would be problematic when students take Regents exams at different grade levels across schools/districts due to sequencing choices (e.g. for math and science) and when all students in the same class may not take the Regents exam.

1c – Developing New NYS Assessments for All Currently Non-Tested Content and Subject Areas

The state could consider developing new state assessments for all currently non-tested content or subject areas. The Task Force accepts the development of new state assessments as a possible long-term option.

The Task Force further recommends that all new assessments for all content areas become some combination of pen/pencil, computer-accessed, or performance assessments as appropriate and required for the subject. This would apply to all subjects including core content subjects, music, art, physical education, and career and technical education (CTE).

For any grade and content area identified by the state as a candidate for a new assessment, the state would follow current procedures to develop and administer the new assessment. This approach may apply to teachers of all subjects and content areas currently not assessed at the state level.

Task Force members considered this approach to be consistent with NYS law, would supply valid and reliable content-specific data aligned with NYS Standards, and that NYS assessments provide multiple years of existing data with predictive power. Some members considered that new assessments could add to the rigor and validity of the NYS assessment portfolio and that new testing formats have the potential to measure more sophisticated student learning. The number of new assessments that would be needed may be staggering and may require a phase-in schedule.

Also, the Task Force realizes that the development of such an array of new assessments is very resource-intensive in terms of time, people, and money. Concerns were raised that NYS would create an “over-testing” program in light of current research that suggests over-testing does not raise student interest and enthusiasm for learning and does not raise higher-order thinking skills.

2. Allow Districts to Choose from a Menu of Options for Assessing Student Growth

2a - District-Developed Assessments

Wherever state exams do not exist, districts could elect to develop pre-tests, post-tests, or both pre- and post-tests as needed. The Task Force accepts District-Developed Assessments as an option. The Task Force also suggests that the Board of Regents consider allowing for the District-Developed Assessment to supersede a state exam, under certain unique circumstances.

Teachers and administrators within a district who share common knowledge and experience in a subject area would design assessments where a gap exists or where the district determines that a district-developed assessment would yield the best results to measure student growth for that particular subject or course. This approach may apply to all teachers if a district-developed assessment is determined to be the very best choice for a pre-test, post-test, or both pre- and post-test.

Many members of the Task Force valued assessments that could measure content more closely aligned to classroom instruction and recognized the potential to create uniform assessment across the district that could provide timely student performance results. The timeliness of those results would increase the ability to monitor student learning and adjust instruction. The Task Force also recognized the ability to build upon high quality assessments that are already used in the classrooms of their districts and considered that a positive by-product might be increased attention to a more unified curriculum. Because of the need to measure student growth in the context of the wide variance in taught subjects, the Task Force valued the flexibility and local control inherent in District-Developed Assessments.

However, the Task Force realizes that developing many uniform assessments could be very resource intensive to the LEA in terms of additional costs and loss of teacher instructional time. In order to develop assessments that ensure rigor and comparability, NYS might require a certification process that determines that the local assessment

meets a prescribed set of standards or NYS might approve a methodology through which local assessments must be developed. Another issue is the determination of adequate growth for so many different assessments with very limited data to draw upon. Some members expressed concern regarding bias in a teacher-created assessment, the sole purpose of which may be for teacher evaluation; however, others felt that the expectation of collaborative design mitigated bias to some degree.

The Task Force noted the difficulty of developing assessments for all performance courses (art, music) and for all Career and Technical Education courses (cosmetology, auto body). The Task Force noted possible security issues and wondered about the data system capacity to maintain the assessment results at both the local and state level. The Task Force is also aware of the amount of technical assistance and professional development that would be needed to scale this process to the number of courses offered.

2b - Regionally-Developed Assessments (through BOCES or other Regional Consortia)

Wherever state exams do not exist, districts could elect to partner with other districts to develop pre-tests, post-tests, or both pre- and post-tests as needed. The Task Force accepts Regionally-Developed Assessments as an option. The Task Force also suggests that the Board of Regents consider allowing for the Regionally-Developed Assessment to supersede a state exam, under certain unique circumstances.

Teachers and administrators across districts within a BOCES region or through other Regional Consortia who share common knowledge and experience in a subject area would convene with the support of a BOCES or others to design assessments where a gap exists or where the districts determine that a regionally-developed assessment would yield the best results to measure student growth for that particular subject or course. This approach may apply to all teachers if a regionally-developed assessment option is determined to be the very best choice for a pre-test, post-test, or both pre- and post-test.

The Task Force considered that regionally-developed assessments would offer cooperative opportunities for small or rural districts to work with others to create uniform assessments across the region providing timely student performance results that would increase the ability to monitor student learning and adjust instruction. The Task Force recognized the ability to build upon high quality assessments that are already used in the classrooms of their districts and expand the knowledge and skill of teachers beyond the confines of the district and considered that a positive by-product might be increased attention to a more unified curriculum.

Because of the need to measure student growth in the context of the wide variance in taught subjects, the Task Force valued the potential support and guidance of the BOCES or others in this process. The Task Force recognized that BOCES or others may be able to forge opportunities. Superintendents may look at “like” districts and create a larger database to view student growth, and Superintendents may partner with unlike districts for the purposes of developing greater aspirations for district offerings.

However, the Task Force was concerned this option may be an expensive process. Also, the Task Force realizes that this process of developing many uniform assessments may be very resource intensive to the LEA in terms of additional costs and loss of teacher instructional time.

In order to develop assessments that ensure rigor and comparability, NYS might require a certification process that determines that the regionally-developed assessment meets a prescribed set of standards or NYS might approve a methodology through which regional assessments must be developed.

The Task Force noted the difficulty of developing assessments for all performance courses (art, music) and for all Career and Technical Education courses (cosmetology, auto body). The Task Force noted possible security issues and wondered about the data system capacity to maintain the assessment results at both the local and state level. The Task Force is also aware of the amount of technical assistance and professional development that would be needed to scale this process to the number of courses offered.

2c - Commercially Available Assessments Identified and Selected By the District

The Task Force considered the use of already existing commercially developed, research-based assessments that meet criteria for measuring student growth and that districts could access through a vendor. The Task Force accepts with reservations the use of Commercially Available Assessments as an option.

Currently, many districts purchase research-based assessments that measure student achievement in various disciplines. Those that are most well-known and most widely used measure developmental student learning in early literacy: i.e. Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Writing and Reading Assessment Profile (WRAP), and Developmental Reading Assessment (DRA). Others measure student learning at higher grade levels: i.e. TerraNova, Iowa Test of Basic Skills, and Degrees of Reading Power (DRP).

The Task Force acknowledges that valid, uniform, nationally-normed, and content-specific K-3 assessments exist and that many districts have been using them as diagnostic measures of student achievement. In these districts, there is a degree of familiarity with the implementation of the measure and the use of the data generated.

As with all commercially-available products, there is a cost to the district in time, materials, and people. The Task Force was clear that a district must learn more about any assessment before moving forward with purchase and implementation, especially in regard to identifying the design of the assessment as a true student growth measure.

Any commercially-available assessment administered well demands professional development for teachers and administrators and ongoing support for implementation. The Task Force voiced concern that the addition of these assessments may add needless testing time to the school day. The Task Force also realizes that challenges to continuity may exist if a district changes vendors or if students change districts and different districts use different assessments.

2d - Commercially Available Assessments Selected by the District from a Menu of Assessments Identified as Rigorous by the State

The Task Force considered the ability of a district to access an official State Education Department list of vetted and acceptable assessments from which the district may select and purchase a state sanctioned, existing, commercially developed, research-based assessment from a vendor that meets criteria for measuring student growth. The Task Force accepts with reservations the use of a state sanctioned list of commercially available assessments.

Potentially, the State Education Department could vet, identify, and approve a list of commercially available assessments that meet criteria for measuring student growth and can be purchased through a vendor.

The Task Force accepts that there are a number of commercially available assessments that may meet the need of districts to measure student growth. The Task Force appreciates the resources necessary in time, people, and money to identify and vet commercial products as viable solutions to the measure of student growth and recognizes the benefits of using state resources to conduct vetting tasks and eliminate the burden of individual districts to do so.

The Task Force is concerned that an officially sanctioned list may limit the number and types of commercial options available and is concerned that the state may eventually obligate districts to select and purchase from the list.

2e - Group Metric from State Assessments

In circumstances in which multiple teachers function as a team, each teacher could receive a score based on the state assessment result of their collective students. The majority of Task Force members felt that there is strong opposition to the use of a group metric from State Assessments. They considered removing this option from the report; however, a decision was made to retain this section to reflect the extent of the discussion.

A Group Metric from State Assessments might apply, for example, to a group of 6th grade science and social studies teachers in one school who function as a team with 6th grade math and ELA teachers. Each teacher would receive a score based on the student growth results of state tests in math and ELA, administered by the 6th grade math and the 6th grade ELA teacher.

Many Task Force members expressed concern that it would be perceived as unfair if an evaluation depended on the work of another teacher. Members also pointed to the weaknesses of a group metric in informing targeted improvement in classroom instruction. Some Task Force members noted that each educator's evaluation is supposed to have three components, and that the growth component should ideally confirm the evidence of an educator's performance that is gathered from the other two; they felt that because a group metric is less likely to be aligned to the curriculum taught in the classroom, it would not serve this purpose.

Because most elective courses at the high school level are aligned with a department that delivers content and skill-based Regents assessment, and many content and skill standards in electives overlap with Regents courses, some members felt this may be appropriate for some sets of teachers but not others. For example, some felt it may be a reasonable long-term solution for performance courses at the elementary and middle levels. Some also expressed optimism that a group metric may improve interdisciplinary curriculum development and a shared investment of student growth in ELA and math skills.

If applied, a group metric would require regulation that specifically defined the portion of student growth that would be assigned to elective teachers.

2f - District-Wide Growth Goals for Selected Assessment

A district sets broad expectations for students' progress in every grade and subject and provides a rigorous framework for schools to use to set differentiated goals for each school and classroom. Then principals and teachers work together to identify a set of goals related to the growth of the teacher's students. The teacher's measure of effectiveness is based on the degree to which those goals are met by the end of the year. Goals will be set by the teacher and principal or the teacher in partnership with others such as the guidance counselor, special education consultant, library/media specialist, ESL department chair, other department heads, or coordinators. Goal-setting partnerships are decided with principal or supervisor input and approval. The Task Force accepts this option in cases where this is the best possible measure for the intended purpose and the growth goals framework is faithfully carried out.

The Task Force recognizes the value of setting student growth goals in meeting the diverse needs of teachers more closely tied to classroom instruction, that teachers would have greater flexibility to change growth goals from year to year and could focus on varied opportunities for improvement, and that teachers could benefit from others knowledgeable in the field (i.e. BOCES professionals, department heads, guidance counselors). This might also work well in conjunction with performance measures of learning and might be a good fit for AP/IB teachers.

The Task Force is aware that this is time-consuming and takes a lot of resources to create goals for all teachers within a building and then monitor those goals over the course of a year. It may be especially time-consuming if a single teacher teaches many different classes and needs to set goals for each class. The Task Force is cautious that this option could be used to cheat the system by creating non-challenging goals.

2g - Performance Assessments Accepted By Professions

The Task Force considered Performance Assessments Accepted by Professions as indicators that a student has met a standard of proficiency that could be performance or industry determined. The Task Force does not endorse this as an option.

The Task Force considered that music evaluations through NYSSMA or Career and Technical Education (CTE) assessments through National Occupational Competency Testing Institute (NOCTI) might be used to determine student growth measures, that a degree of uniformity may exist and that multiple years of data may be accessible.

However, the majority of the Task Force believes that these assessments were not developed for high-stakes evaluation and would require significant work to get them to

the point that they are appropriate for teacher evaluation, that they are not aligned to NYS Standards, and that there is no corresponding benchmark assessment. Also, these assessments do not differentiate among students; there is no data collection system in place; tests can be coached; high SES-students have greater access to tutors, and the prevalence across NYS varies (presidential Fitness Test has high use; NYSSMA has low use). Also, testing often occurs outside of the school day, and the assessments are often costly to both the student and the school.

Student Growth Measures for Principals

Principals		
	Elementary/Middle (with 4-8 in 2011-2012; all in 2012-2013 and beyond)	High School (2012-2013)
Statewide Measures of Growth 20 percent (25% with approved VA model)	<ul style="list-style-type: none"> ○ Result of student-growth/value-added model for all tested grades/subjects; add grades or subjects as value-added model applies ○ Course grades/failure rates 	<ul style="list-style-type: none"> ○ Result of student-growth or value-added model as applied to English and Math Regents performance; add subjects as value-added model applies <p><i>Near-term measures:</i></p> <ul style="list-style-type: none"> ○ Progress toward graduation rate against predicted metric ○ Credit accumulation ○ Course grades/failure rates
	Not recommended: <ul style="list-style-type: none"> ● Growth within student subgroups ● Student growth in non-tested grades and subjects 	

Near-Term Measures in Addition to Results of Growth/Value-Added Model

Task Force members expressed concern that there may be nothing in the value-added model for 2012-2013 for high school principals (although staff has said that a new ELA and math value-added model at the high school level is possible). The group identified possible “near term” measures that would expand measures for high school principals while other subjects are added to the value-added model:

- A Task Force member acknowledged that *credit accumulation* was used in NYC for high school principals. Another member suggested that further research on how others have used credit accumulation and its reliability as a measure should be done. There was concern about how to translate credit accumulation into a

growth measure. The Task Force member acknowledged that using credit accumulation as a measure allowed for comparison across cohorts.

- NYC described its current approach to including Regents exam performance in principal evaluations, as well as the number of Regents courses a student earns credit for. This could be an option for the state.
- A Task Force member recommended using *course grades* as a reliable predictor of student performance in middle and high school, as well as a stable predictor of college GPA, and suggested that further analysis be done on the research that has been done on this. The member recognized course grades as a no-cost measure that covers all subjects in all grades. Task Force members thought this possibility warranted further discussion.
- A Task Force member discussed *graduation rates* as a measure that warranted further discussion, indicating that high school principals want parity between elementary and middle school principals, as well as teachers, for this measure since the measure is typically a district-wide goal. One member indicated that NYC takes into consideration the characteristics of the incoming ninth-grade class and determines it as a baseline for which principals are accountable. There are also peer groups of schools identified to level the playing field. He also indicated that specialized high schools are most unhappy with growth, while SURR schools sometimes perform better on the growth piece. Another member voiced concern that principals in high-performing schools should neither get penalized if they're already almost at the ceiling for graduation rates, nor should they have a distinct advantage over colleagues with struggling students.

Growth Within Student Subgroups

Some Task Force members did not think that the results of the growth/VA model for particular student subgroups (students with disabilities, English language learners, highest/lowest achievers) was an appropriate growth measure for either elementary/middle or high school principals. One member suggested that growth of student subgroups could be considered in the principals' 60% other measures as part of a principal's school-wide goals.

Student Growth in Non-Tested Subjects

Task Force members did not think student growth in non-tested grades and subjects was an appropriate measure for elementary/middle or high school principals. Members acknowledged that it is important to see student growth across *all* areas to ensure breadth. But some felt the measures that would have to be used for teachers in some

non-tested grades may be less rigorous and should not therefore be included as measures of growth for principal evaluation. Questions were also raised regarding how results from different assessments in different subjects and grades could be aggregated and how those results would be weighed (for example, by discipline).

Locally-Selected Measures of Student Achievement

We, the appointed members of this Task Force, are pleased to present our recommendations to the Regents.

Our Task Force has taken this work very seriously. We have engaged in thoughtful and often challenging dialogue to reach consensus on numerous topics. We represent various and often opposing perspectives on issues of bargaining and practice, which we believe enhanced the dialogue of our Task Force. With few exceptions, where legal interpretations are necessary, we have reached consensus on what we believe to be fair, objective, and manageable recommendations for eventual regulations and guidance to the Commissioner.

Mindful of the delicate balance necessitated by the spirit and intent of the law; the integrity and professionalism of public school educators; the tradition of local school district control; and the aspiration to achieve the highest levels of student achievement on state, national, and international academic standards, this Task Force implores the Commissioner and the Board of Regents to preserve the “authenticity” of assessments for this 20% (15% in subsequent years) of a teacher’s and a principal’s evaluation.

In an era when states are being swept up by the wave of high stakes standardized measures and political attempts to categorize and rank professional work, we believe strongly that authentic student work that illustrates critical thinking, problem solving, and deep thought and work can be preserved as a measure to respond to the questions: “Are children learning from and with this teacher?” and, “Are children learning in this school as a result of the presence of this educational leader, the principal?”

Unlike other states, New York can stand out as a leader to say, “Authenticity matters.” “Rigor matters.” With the use of State Academic Standards; a clear and ambitious definition and rubric of rigor to deem assessments rigorous in every school district; and a means of decision-making that is manageable and professional, we believe local assessments can be just that— local— while achieving the intent of the law and the spirit of local school district culture.

We believe choice in locally selecting such assessments is a noble effort for the State of New York and that it is “do-able” even in a high-stakes climate.

Most importantly, in order to preserve the integrity and quality of every classroom and setting where children learn in school districts across the state, children must not be

subjected beyond the tipping point of excessive test-driven curriculum and test-taking routines. The use of locally selected assessments from a menu of options for locally developed or vendor developed authentic assessments of learning meets the intent of this law while preserving school as school should be and learning as learning should be for all learners, or as we call it, “School the way it ought to be!”

This document moves the conversation from theory to practice and from law to recommendations for use in drafting regulations. It illustrates that which we believe complies with law and implements sound professional practice. It is not perfect. While keeping the bar high, it respects that no two districts are alike in population, geography, size, capacity, and economics with bona fide and comparable options for all.

On that note, we would hope that as this component of evaluation is phased in for some teachers in 2011-2012 and all teachers in all subjects in 2012-2013 and beyond, that the invitation to take risks remains high for teachers wishing to press themselves and their students with rigorous assessments while the stakes remain low until proverbial bugs are out of the new system.

Recommended

The following table summarizes areas where the Task Force has reached consensus on recommendations.

Both Principals and Teachers	
This component of the evaluation system should be based on assessments that are not standardized State tests.	The group feels very strongly that state tests should not be a part of the 20% locally selected measures. This would place too much emphasis on the State tests and have a profoundly negative impact on instruction. Further, many group members feel that the use of state tests for the local options is in contrast with the language of the law. This part of the evaluation should provide teachers and principals an opportunity to show achievement beyond standardized tests.
These locally selected measures must add value to classroom instruction.	There needs to be a purpose/use other than solely for evaluation that is not undermined by its use as part of evaluation.
The law requires multiple measures of achievement.	More than one assessment is required. These can be discrete, one time measures or may be growth measures over multiple points in time.

Teachers ELA/Math 4-8

All options must be verified for comparability, validity, rigor, degree to which the assessment can aligned to State standards, and degree to which the assessment(s) match classroom instruction.

General Recommendations	Comments
<p>The 20% must include multiple assessments. The term “assessment(s)” pertains not only to “tests,” but may include other forms of assessment that relate to student achievement. These assessments can measure growth over one or more points in time or can measure achievement at a single point in time.</p>	<p>Multiple-measure systems improve the accuracy and stability of teachers’ evaluations by reducing reliance on any single measure of a teacher’s performance. This local option should be aligned with NY State Common Core Standards, meet statewide criteria, and consist of multiple measures of student performance such as:</p> <ul style="list-style-type: none"> • Criterion referenced test • Curriculum based assessments • Formative assessments • Norm referenced tests • Performance assessments • Portfolio or student work • Summative assessments
Teacher Options	Comments
<p>1. Pre-approved menu of acceptable 3rd party assessments</p>	<p>The Task Force agrees that this is one viable option for a locally selected assessment. There is also the risk of adding more high-stakes <u>tests</u> (vs. assessments) and potential unintended consequences.</p>
<p>2. Districts issue RFP for assessment and verify comparability and rigor.</p>	<p>The Task Force agrees that this is one viable option for a locally selected assessment. This would require an RFP process and has fiscal implications. These assessments would be standardized across the district. There are the same potential risks as with option 1.</p>
<p>3. Districts develop their own assessments and verify comparability and rigor</p>	<p>The Task Force agrees that this is one viable option for a locally selected assessment. This option is resource-intensive but offers opportunities for professional development and building local capacity. These assessments would be standardized across the district.</p>
<p>4. School-wide, group or team assessments (group metric)</p>	<p>The Task Force thinks that this is one viable option for a locally selected assessment but not likely for ELA and Math 4-8. A district may decide to use a school-wide or grade-wide assessment to emphasize team-based teaching or address a district/school goal.</p>
<p>5. Districts allow schools to utilize teacher-created assessments</p>	<p>The Task Force supports teacher created assessments with a mechanism in place to verify comparability and rigor (a rubric tool). Subjectivity in the system cannot be fully controlled for. This option encourages collaboration with the principal and professional development.</p>

Principals	
All assessments must pass muster for rigor (by rubric in standard) and comparability across schools in a district.	
General Recommendations	Comments
Based on the local assessments / multiple measures.	Does not need to be an aggregate but could have focus on a district priority such as ELLs or writing. There must be a balance between the big picture and focused goals. Some goals may span multiple years.
The scoring would be done locally.	The evaluator of principals (presumably superintendents) provides the final rating to the SED, not the data that the SED then feeds into a formula.
Options	Comments
Locally selected assessment that illustrates the achievement level of ONE identified group of students (ELL, a grade level, Special Education, Free and Reduced Lunch, or other) on ALL or MANY content/subject areas. and/or Locally selected assessment(s) that illustrate the achievement level of ALL students on ONE identified content/subject area.	The assessment(s) could be the same as the ones used for teachers of the identified students or an alternative assessment for this purpose. The assessment(s) could be a single test of achievement, growth from one year to the next, or pre/post improvement.

Further Discussion Required

This table includes options that may be viable that have been identified by the Task Force but have not been discussed.

Teachers
The Task Force has had limited discussion of locally-selected measures for teachers in non-tested areas.

Principals	
Options	Comments
There has been limited discussion regarding options for principals.	
Other assessment data such as graduation rate, drop-out rate, AP, IB other Regents equivalents take and pass rates, and other compilations of achievement that are not based on standardized tests.	The Task Force needs further discussion on whether these could be considered a “measure of achievement” but believes these might more appropriately belong in the 60% component of the evaluation.
Selected metrics from local assessments used for teachers	The Task Force needs further discussion on this option.

Not Recommended

The following were suggested to the group as possible options. These options were not created by the Task Force. *The Task Force feels strongly that these options should not be recommended. Please see Appendix F—Workgroup Opinion on State Tests for 20% Local for the Task Force’s position on why state tests should not be an option for the locally selected assessments. For an illustration of possible alternative scenarios, see Appendix G—Scenarios for Use of State Test Results in the Other 80%, by member of Metrics Workgroup, March 24, 2011.*

Teachers and Principals

- Utilization of state tests for any locally selected measure was rejected.

Teachers

- School-wide, group or team results based on state assessments.
- District-wide goal setting process for use by principals/ teachers with any state assessment they select.

Principals

- Achievement on state tests (% proficient)
- Performance of student subgroups (SWD, ELL, highest or lowest achievers) on single subject achievement.
- Aggregate achievement on locally selected measures for all students, all teachers.

Task Force Discussions

Issue 1: Local comparability

2.3. (ii) twenty percent shall be based on other locally selected measures of student achievement that are determined to be rigorous and comparable across classrooms in accordance with the regulations of the commissioner and as are developed locally in a manner consistent with procedures negotiated pursuant to the requirements of article fourteen of the civil service law.

Task Force Positions

For the purposes of the local assessment option, the Task Force began with the assumption that comparable means that a single district-wide assessment system will be used to gauge effectiveness for all teachers of the same subject and grade within that district. Note: The term “assessment system” needs to be clarified, particularly with regard to how an assessment system addresses comparability.

All Task Force members agreed that district-wide comparability measures could include teacher-created assessments, so long as common standards of comparability and rigor are met. As stated by one Task Force member:

“The New York State Education Department would need to craft a rubric that would measure the degree to which an assessment is rigorous, valid, reliable, and meets core standards. This rubric would be used for all assessments, allowing for comparability in each of the categories. The law does not currently specify that comparability means the same assessment, leaving open the possibility that different assessments could be comparable if they achieve the same score using the SED rubric... Subjectivity may be a factor when using the SED-created rubric to evaluate the teacher-created assessment. However, if more than one person were to independently evaluate an assessment and compare results, such as an assistant principal, certified consultant teacher or a department representative along with the evaluator, the chances of unintended subjectivity would lessen and give evaluators more confidence that their determination is corroborated by other, trained and experienced individuals. “

Issue 2: Rigor

Task Force positions

The Task Force has been working on defining rigor. No consensus has been reached on an appropriate, comprehensive definition of rigor. There was unanimous agreement with the assumption that the local assessment should be aligned to the NYS Common Core standards.

Some questions surrounding rigor are reflected in this discussion that was posted to Hope Street

“The issue on the table, which [we are] working to define, is how to define “rigorous”; as in “locally selected measures that are rigorous and comparable across classrooms.” It can be assumed that a menu of vendor-made assessments that SED approves for local use (i.e. the DIBELS Assessments) will have been vetted for their rigor. Locally developed assessments should conform to SED expectations that are clearly outlined by checklist or rubric. The question of how the assessments are developed by localities is being left to the collective bargaining agreements of each district. “

- ❖ One Task Force member offered a “Rigor Checklist” created by the University Of North Carolina Of Greensboro:

Rigor Quick Check

Reminder: Look at what students are being asked to do.

Content:

Is the content part of the state standards?

Does it include basic skills and important concepts?

Does the content require students to apply core academic knowledge to problems or issues?

Instruction:

Does the instruction require students to engage in higher order thinking skills?

Are students required to engage in elaborated communication?

Do they have to explain or justify their conclusions or thinking?

Assessment:

Is the assessment aligned to the lesson goals?

Does the assessment measure SCOS content?

Do the students have to use higher order thinking skills on the assessment?

Do students have to explain or justify their conclusions or thinking?

❖ Another Task Force member offered this discussion on rigor:

Locally developed and/or locally selected assessments and assessment processes are characterized as:

- *Content-Rich: This is based on a template to determine rigor and indicators of higher-order thinking.*
- *Rigorous: This is based on a template to determine rigor and indicators of higher-order thinking.*
- *Curriculum Aligned: Alignment determined by comparison with “Common Core” are content standards*
- *Assessed Authentically: Standards-based processes are developed and utilized by teachers or contracted professionals*
- *Reliable: Assessments are determined to be consistent across classrooms and overtime through rubric use and analysis*
- *Valid: Assessments are determined to measure what is intended to be measured through standards alignment analysis and other measures*
- *Professional Development: Teachers or contracted professionals are trained in authentic assessment, academic rigor, and specific use of selected assessment tools.*
- *Inter-rater reliability: Those scoring are trained and deemed proficient in scoring.*

Issue 3: Individually Teacher-Created Assessments

Task Force positions

The Task Force has had numerous discussions surrounding this option. The group has reached agreement that it did not want to preclude this option.

One group member wrote:

Pros:

Allows teachers to:

- *customize assessments to match the unique make-up of their student populations, such as ESL or SWD cohorts.*
- *craft creative and higher-order thinking assessments beyond the standardized test format- such as performance-based assessments.*

- address district or building-specific learning goals.
- effectively measure the degree to which students have mastered what is actually being taught.
- use relevant and meaningful data to direct curriculum and instructional methods.
- supplement the material assessed on standardized tests to decrease the likelihood of test-centered curriculums.
- gain professional development. Working with evaluators to ensure that assessments meet NYS standards will provide teachers with valuable on-the-job experience, particularly for newly hired staff members.
- feel more empowered with regard to the student achievement piece of their evaluation. Increased buy-in to the new evaluation system might result if such a choice was provided.
- have greater opportunities for collaboration and communication with evaluators. Evaluator feedback would inform the instructor and, in turn, afford the evaluator greater insight into what is actually being taught and assessed in the classroom.
- update the curriculum and local assessment tool as needed. (e.g. - An ELA teacher might wish to use new literature, possibly necessitating a change in the local assessment. If the assessment were teacher-created, such a change would be easy to make.)

Neutral Comments:

- Many schools already require that teachers submit their final course examination to the administrator. If the 20% Teacher-Created Local Assessment was based on student achievement, then such administrators would be evaluating assessments they've already seen or are used to seeing in the course of their regular duties. In other words, this would not be additional work for administrators that already preview the final exams given in their schools.
- As the Teacher-Created Option would be one of several options on the menu, schools would not be under any obligation to choose it. However, by allowing the Teacher-Created Option on the menu, schools would have the opportunity to take advantage of the above listed benefits should it be locally negotiated and determined to meet the needs of the district.

Cons:

- Additional time would be required of administrators to evaluate Teacher-Created Assessments. Administrators have ever increasing time constraints and responsibilities, including their new Principal Evaluation requirements. One way to aid over-worked administrators would be to permit assistant principals or teacher leaders such as certified consultant teachers or department representatives to evaluate the assessment

tool using the SED rubric. The administrator might even have a list of struggling teachers he or she needs to look at personally in terms of their assessments- but permit other, highly effective teachers to submit their assessments to teacher leaders. As not all districts would be comfortable with such arrangements, this possibility might be a locally negotiated item.

Another way to lessen the time burden would be to have the administrator evaluate the assessment tool during the pre-evaluation meeting with the teacher. As most districts require such a meeting prior to the first classroom observation, the extra time to look over the Teacher-Created Assessment would not be quite as onerous as scheduling a separate meeting.

It is noteworthy, however, that all educators- be they administrators or instructors- are being asked to take on additional responsibilities in the interest of providing quality teacher and principal evaluations- an important step in improving student achievement...

- Teachers might need release time to create a local assessment. However, an individual teacher might require less release time, if at all, than a group. Creating assessments throughout the year is part of an instructor's responsibilities. Many may already have exams or assessment tasks in place that would meet the appropriate standards. The need for release time could be locally negotiated as is necessary."

Another group member wrote:

"This is an area of critical significance and poses a myriad of concerns, many of which you captured here. As we continue to look deeper at the possibility of using teacher-created assessments, the concern of whether we truly could ever ensure comparability across classrooms remains the single most important question to answer. Being well grounded "in the trenches", you and I both well understand the implications of how teacher-created assessments could play out in ways we would never want the profession to proceed. The driving question here is now that we are using assessments as part of evaluating the effectiveness of teachers and principals, how can we not compromise the reason we give assessments in the first place - to inform and adjust our instruction to better meet the needs of all of our students."

There were some group members who felt that this option would not be feasible, especially in terms of the principal's responsibility for monitoring all of these assessments.

Issue 4: Scope of Collective Bargaining

Task Force positions

The Task Force had several discussions surrounding this issue, but eventually reached consensus that absent clear and direct guidance in the language of the authorizing law, we would not attempt to define the parameters of what should be included in collective bargaining at the local level with regard to locally created assessments.

While we as a Task Force have not reached agreement on the scope of bargaining as articulated in the law, we concur that local voices must be at a table, be it bargaining or local committee or both, to match the professional work in a district with the academic standards of the state and expectations for rigorous assessment for the purpose of professional evaluation.

A memorandum was prepared by one Task Force member to reflect, in part, the depth of the discussion around this issue (See Appendix I—Local Bargaining Issues, by one member of Workgroup on Locally-Selected Measures).

List of Appendices

For additional context on various issues, see:

Appendix E—Synthesis Document of the Locally-Selected Measures Workgroup from January 2011 Meeting

Appendix F—Workgroup Opinion on State Tests for 20% Local

Appendix G—Scenarios for Use of State Test Results in the Other 80%, by member of Metrics Workgroup, March 24, 2011

Appendix H—Teacher-Created Assessments, by member of Workgroup on Locally-Selected Measures, Jan 30, 2011

Appendix I—Local Bargaining Issues, by one member of Workgroup on Locally-Selected Measures

Other Measures of Effectiveness: Teachers (60%)

Standards	There is consensus that the evaluation system should be grounded in the New York State Teaching Standards.
Choice of rubrics to assess proficiency against standards	The regulations and guidance on teacher evaluation should reflect flexibility in the choice of rubrics to assess teacher proficiency. There is consensus around the State developing a list of pre-approved rubrics from which a district could choose, as long as there is a variance process allowing districts flexibility to use a pre-existing rubric currently used in the district.
Required elements	<ul style="list-style-type: none"> • Multiple Sources of Feedback: There was consensus that multiple measures may be included (e.g., multiple observers, additional options negotiated locally, consistent with collective bargaining). • Observation by principal or other administrator is required.
District options	<p>In addition to the requirements above, the Task Force reached consensus (except as noted) that the following options can be considered, subject to local bargaining agreements. This list is meant as a starting point and should not be thought of as all-inclusive.</p> <ul style="list-style-type: none"> • Classroom Observation: <ul style="list-style-type: none"> ○ Observation by trained in-school peer reviewers ○ Observation by independent, trained observers (<i>mixed views</i>) • Video-taped classroom sessions (<i>mixed agreement: more support for use for professional development, personal growth/reflection, "norming" evaluators than as a method of observing for evaluations</i>) • Structured review of student work • Teacher artifacts using "portfolio" or "evidence binder" processes • Feedback from students, parents, and/or other teachers using structured survey tools • Individual professional growth plan: • Teacher self-reflection
Training	There was consensus that all evaluators should be trained
Professional Development	Individualized professional development plans should include professional growth goals that are individually established by both the teacher and the principal or other appropriate personnel

The Teaching Standards

The Task Force's discussion of other measures began with the integration of the NYS Teaching Standards, including the degree to which these standards will be embedded in the 60% other measures. There is broad consensus that the Teaching Standards should serve as the foundation of the 60% other measures.

- Members of the Task Force feel strongly that the Standards must be an integral part of the 60%.
- Members reached consensus that a school must address all seven standards in a year but not necessarily all the elements each year.

Proposed Criteria for Teacher Practices/ Rubrics

- Teaching practice rubrics must cover broadly the New York state standards or, if rubric is only designed for classroom observation, those standards that relate to classroom practice.
- Rubrics that have proven correlation to student learning outcomes based on high quality academic research are strongly preferred if available. Flexibility is needed to allow for the ongoing acquisition of knowledge in this area.
- Rubrics describe multiple levels of effectiveness for each standard or element. If a rubric does not have four levels that match the NYS ratings of Highly Effective, Effective, Developing and Ineffective, the rubric's summary ratings must be easily convertible to the four levels NYS has adopted.
- Expectations for each performance level must be clearly defined in the rubric and differentiated from each other. The differences among performance levels must be based on meaningful and observable differences in teacher practice or student learning.
- The rubric must use clear and precise language that facilitates common understanding among teachers and administrators.
- To the extent possible the rubric should rely on specific observable behaviors in the classroom with direct evidence of student engagement or learning in addition to evidence of teacher actions or inputs.
- Rubrics must be specifically designed to assess the effectiveness of teachers.

- Rubrics must be accompanied by descriptions of any specific training and implementation details that are required for the rubric to be effective.
- Rubrics may be applicable to all grades and subjects or could be designed explicitly for specific grades and/or subjects, but if so, they will only be approved for use in the grades or subjects for which they are designed.

Assessment Approaches for Teachers

There was consensus that all of the following options be part of the collective bargaining process. Options considered by the Task Force include, but are not limited to:

- Observation by trained in-school peer reviewers
- Observation using video of classroom practice and any approved evaluator
- Video-taped classroom sessions (mixed agreement: use for professional development, personal growth/reflection, “norming” evaluators)
- Structured review of student work
- Teacher artifacts using “portfolio” or “evidence binder” processes
- Feedback from students, parents, and/or other teachers using structured survey tools
- Individual professional growth plan
- Teacher self-reflection
- Professional growth goals that are individually established by both the teacher and the principal or other appropriate personnel
- Participation on committees and work at the State Education Department
- Participation in state professional organization leadership roles

Classroom Observations

Task Force members have reached consensus that observations should be a part of the 60% other measures and that there should not be a set number of minimum observations.

Who is the observer?

Task Force members reached consensus that at least part of the 60% should be based on observation of teacher classroom practice by *trained principals and administrators*, subject to collective bargaining agreements.

The consensus is that *peer review* of teacher performance should be permissible (not mandated by the state), subject to collective bargaining.

There is a split regarding whether the use of *outside evaluators* should be permissible, subject to collective bargaining. Some believe that outside observers would be unfamiliar with the culture of the school and the issues endemic to a particular school or district.

The Use of Rubrics

Rubrics and Academic Research

The Task Force had concerns regarding whether a menu of rubrics approved by the state or locally developed rubrics meeting state criteria should be required to have research-based evidence of results. While the consensus of the group is that it is preferable to have rubrics about which academic research has found a correlation to student outcomes this standard might exclude rubrics, existing or future, that may have a valuable place in teacher evaluation.

Rubric Development

There was much discussion regarding whether there should be one State-developed rubric that all districts would be required to use, whether the State should develop a list of approved rubrics that districts could choose from, or whether the districts could be free to choose any rubric or develop their own. There is consensus that:

- The State should develop criteria for the selection or the development of local rubrics.
- The State should develop a list of pre-approved rubrics from which a district could choose.
- The State should establish a variance process for approving individual districts' use of a rubric that is not on the pre-approved list.

Rubric Utilization

There is consensus that it is necessary for every district to choose a rubric and that educators and evaluators must be trained in how to use it. As the use of evaluation rubrics evolves it will be important for both teachers and administrators to stay current with the research being done on their use and on how to incorporate the findings of such research to improve teaching practices and student learning.

Other Measures of Effectiveness: Principals (60%)

Introduction

New York State statute provides that beginning in July 2011 a new evaluation system will measure the performance of teachers with grades 4 through 8 English language arts and math assignments, along with their respective principals. This new evaluation system will go into effect for the remaining teachers and principals beginning July 2012. Sixty percent of a principal's professional performance will, in some way, broadly assess principal leadership and management actions. The purpose of this report is to recommend a framework of standards, rubric options, required evaluation elements and measures, sources of evidence, and weighting considerations to the New York State Board of Regents and the Commissioner of Education.

State planning for principal evaluation began in 2008, with funding from The Wallace Foundation. The funding's broad purpose was to enable the state to create a Cohesive Leadership System (Augustine et al., 2009¹) by aligning its standards, policies, and funding for

- leadership preparation
- leadership licensure
- leadership development
- principal evaluation

The state's work was guided by a director, national expert Joe Murphy, and consultants from Education Counsel. The state formed a statewide advisory committee to guide this work, with representatives from two state principal associations (CSA and SANYSS); state superintendent association (NYSCOSS); leadership preparation (MCEAP and CADEA); state school boards association (NYSSBA); and district officials from NYC, Rochester, Buffalo and the mid-Hudson region. The committee met monthly over an 18-month period and was supported by a series of subgroups that focused in-depth on two policy areas—preparation and evaluation. As a result of the state education staff and committee work, the state:

- adopted the ISLLC standards to frame leadership policies;

¹ Augustine, C. H., Gonzalez, G., Ikemoto, G. S., Russell, J., Zellman, G. L., Constant, L., et al. (2009). *Improving school leadership: The promise of cohesive leadership systems*. Santa Monica, CA: Rand.

- used the standards and best practices research to design model guidelines for leadership preparation programs, for a competitive grants process;
- used the ISLLC standards to revamp the state leadership assessment for licensure;
- used the standards and best practices research to design model guidelines for leadership academies, funded through the grant; and
- used the standards, best practices research, and state models to design a principal evaluation model.

The proposed principal evaluation model was presented to the NYS Regents in June 2010, prior to the initiation of the NYS Task Force on Teacher and Principal Effectiveness. It included design principles for effective principal evaluation, stressed its developmental purpose, and included key elements on

- personal professional goal setting,
- school wide goal setting,
- alignment to the ISLLC standards as a benchmark of performance,
- grounding in local student improvement needs and priorities,
- feedback from multiple sources,
- evidence and reflection of principal growth and development, and
- supervisor support.

When New York State was awarded Race to the Top funds in late 2010, the work of that advisory group was carried over to the newly formed Regents Task Force on Teacher and Principal Performance and served as the foundation for furthered discussions and explorations of a principal performance accountability system.

Overview of Recommendations

Standards	There is broad but not complete agreement that the evaluation system should be grounded in the 2008 ISLLC standards
Choice of Rubrics to Assess Proficiency	The regulations and guidance on principal evaluation should reflect flexibility in the choice of rubrics to assess the data on principal proficiency. There is mixed reaction to whether this should be a menu or free choice.

Required Elements of a Principal Performance Evaluation	<p>A principal performance evaluation should</p> <ul style="list-style-type: none"> • broadly assess principal leadership and management actions; • occur annually; • yield results that allow for the differentiation of principal performance and have the capacity to differentiate between the four HEDI titles; • promote collaboration, ongoing communication, timely feedback, and trust between the supervisor and the principal being evaluated; • be based on multiple measures that are state defined; • include feedback from the supervisor plus two additional stakeholders; • include at least two sources of evidence for each measure, with feedback being one source of evidence; • allow flexibility for districts to take into consideration factors such as a principal’s tenure status, years of experience, years serving in a particular school, and local context; and • be achieved using a standardized rubric, in which all evaluation tools and data collection sources are – to the extent practicable – research-based and psychometrically strong, with demonstrable validity and reliability.
Required Measures of Principal Performance	<p>The assessment should include measures of the following:</p> <ul style="list-style-type: none"> • Broad assessment of principal leadership and management actions • Assessment of principal’s skill in supervising and developing teachers • Progress against individual professional growth goals • Progress against ambitious and measurable school academic or learning environment goals
Sources of Evidence	<p>There should be multiple sources of evidence for each measure. The following optional sources are suggested:</p> <ul style="list-style-type: none"> • School visits • Formalized and standardized feedback from students, parents, teachers, community stakeholders • Principal self-reflection • Review of school documents, systems, and records • Results and findings from existing state accountability processes • Other sources as determined locally

Weighting of Evidence

- Evidence should be scored holistically, with local districts determining the weight to be given to each element.
- The regulations should not mandate specific weights or numbers/ranges of points for particular measures within the 60%; districts should be free to determine how the measures within the 60% are scored.

Standards for Principal Evaluations

The Task Force indicated broad agreement that the evaluation system should be grounded in the 2008 ISLLC standards. Three different ideas about which standards should be used were discussed at length. These discussions centered on

- adopting ISLLC standards statewide,
- allowing districts to apply for a variance, or
- developing state standards.

Adopt ISLLC Standards

The Task Force indicated broad agreement that the evaluation system should be grounded in the 2008 ISLLC standards. To begin, many Task Force members identified the standards as the appropriate framework for a principal evaluation in order to “facilitate consistency and uniformity throughout the state in the practice of school leadership.” This desire for consistency was echoed by Task Force members’ use of words such “coherence,” “alignment,” “common language,” and “articulation with other policies” when arguing for the adoption of the ISLLC standards.

Several Task Force members argued for coherence in state policy on the use of standards, pointing out that through the state work on developing a Cohesive Leadership System (through Wallace Foundation support) the state had adopted the ISLLC standards and required it to be the standards to frame funding for leadership preparation programs and leadership academies. The state is piloting new performance assessments for school leader licensure; these assessments are based on the ISLLC standards. Finally, all leadership preparation programs in the state are required to have national accreditation. Most will be using NCATE, which uses the ISLLC standards. Thus, from preparation, licensure and professional development, the state has aligned its leadership expectations around the ISLLC standards.

Further, Task Force members identified Rhode Island, Rochester, and NYC (School Leader Competencies) as having models grounded in ISLLC standards. Additionally, other Task Force members felt strongly that adopting ISLLC standards would allow for

consideration of the full lifecycle of a principal and believed that having more than one set of standards meant there would no real, consistent meaning for an individual. Finally, Task Force members voiced their support for the standards because they are tied to improving student learning and achievement. Consequently, these leadership standards and functions are a framework for focusing on essential areas of principal behavior – what a principal *does* to improve student learning and achievement.

Allow Districts to Apply for a Variance

One Task Force member proposed that districts be granted flexibility in adopting ISLLC standards if the district is already using a proven and effective research-based competency model that has well-developed practice in place: “The Quality Review is a comprehensive, well-established, proven analogue to the ISLLC standards. It was designed by looking at ISLLC and a variety of other rubrics and measures of school and principal performance, and incorporating the best from each. [It] meets or exceeds nearly all the standards set by ISLLC [and] has been in effect for five years, has been improved each year, including the greater focus this year on bringing school curricula and assessments into line with the Common Core standards that the state recently adopted, and therefore makes it a model that other districts in New York State might want to consider adopting.”

A few Task Force members expressed concern about allowing for a variance process around standards. One member noted, “Each potential evaluation system would have to demonstrate comparability in scope, depth, alignment, definitions, and evidence. A formal review process would consequently need to be implemented to ensure this comparability, and such a process is currently not articulated.”

One Task Force member pointed out that there is only one set of teacher standards for the state, without need for local variance.

Develop State Standards

One Task Force member suggested that New York State should develop one set of standards and one evaluation tool and that principals would benefit from a solid and concrete foundation similar to that being recommended for teachers through the use of the teaching standards. Further, the member argued that the Task Force’s proposed framework, grounded in ISLLC standards and designed in concert with materials from other states, would not only provide a sound and consistent foundation for the evaluation of building principals but also would facilitate the development of an evaluation rubric that can more easily correlate to the four bands required by the law.

Rubrics to Assess Proficiency

The Task Force agreed to flexible choice on rubrics to assess the data on principal proficiency, and indicated that regulations and guidance must reflect this need for flexibility at the local level. There was mixed reaction, however, to whether this should be a menu or free choice. The three policy considerations are as follows:

- Provide a menu of state-approved choices and a district variance process.
- Allow for the free choice of any rubric.
- Identify a single statewide rubric.

Provide a Menu of State-Approved Choices and a District Variance Process

Task Force members identified several reasons for a menu choice and variance process. To begin, choice was identified as “very important, as each district is different within the state,” and choice was considered beneficial when using an evaluation system that can “reflect the reality of the principalship in a given district/school.” Further, a Task Force member identified that Rhode Island allows local districts the option to create their own rubric and use rigor, alignment, and standards to qualify. Additionally, the member noted, “Rochester’s evaluation tool is a modification of ISLLC standards. New York City’s touches lightly on standards 4, 5, and 6, and is heavy on 2 and 3.”

Some Task Force members are concerned with “unfunded mandates and [the fact that] some instruments [on a menu] may be costly,” and it was argued that districts should have a choice of assessments because “if districts can’t afford to implement them properly, they are worthless.”

Some Task Force members identified challenges to offering a menu and a variance process. To begin, some believe that variability could undermine the statute’s intent to arrive at statewide uniformity. There are further comparability concerns: “Two administrators doing the same quality work in two districts could have two very different ratings because of district priorities.” Another member noted, “Local districts will need to submit plans to the state on their types of data, rubrics, weighting and calculating scoring, and their alignment to the potential state-mandated performance categories,” and questioned if the state had devised a quality assurance system. Another Task Force member stated, “Discretion regarding the weighting of factors will result in a situation where school districts would pick and choose which standards would be applicable in their locality or, at a minimum, give different emphasis to the various standards. Such an outcome would prevent the possibility of any type of statewide consistency regarding school leadership practice.” Finally, one Task Force member

noted that “many existing evaluation instruments – principal rating scales – have modest reliability, demonstrating the infancy of this field.”

Allow for the Free Choice of Any Rubric

There were, however, challenges to allowing for the free choice of any rubric. One member noted, “Under the law, scoring calculations are to be determined in accordance with commissioner’s regulations. Therefore, local flexibility would be inconsistent with the law.” Other members stated that granting discretion to districts to weight certain factors takes away the standardization of the process and could prevent any type of statewide consistency regarding school leadership practice. One member summed it up, “There should be some rational agreement about the relative importance of the various barometers of principal effectiveness.”

Identify a Single Statewide Rubric

The Task Force had broad agreement that there should *not* be one single mandated rubric or set of assessment tools. One Task Force member felt that “regulations should not rule out effective instruments that could be used through a variance process. There is value and efficiency in trying to regionalize an assessment rubric. If a lot of districts want to use one model because it works for that region, a variance process is necessary.” Other Task Force members cited the local factors, issues, and challenges districts face and the impracticality of identifying one evaluation to assess principal performance. There is noted concern among Task Force members that an evaluation system has to be flexible. One member voiced the concern, “Administrators need to be evaluated based upon criteria which is fair and which is within their control.” Another Task Force member echoed the need for flexibility when evaluating a principal when he stated, “In large urban districts, principals often have no control over [for example] who they hire.”

Required Elements of a Principal Performance Evaluation

The Task Force agreed that a principal evaluation include required elements, which are as follows:

- A principal evaluation should be a broad assessment of principal leadership and management.
- The evaluation should occur annually.
- The evaluation should yield results that allow for the differentiation of principal performance and have the capacity to differentiate between the four HEDI titles.
- The evaluation must promote collaboration, ongoing communication, timely feedback, and trust between the supervisor and the principal being evaluated.

- The evaluation should be based on multiple measures that are state defined.
- The broad assessment should include feedback from the supervisor plus two additional stakeholders.
- The evaluation should include at least two sources of evidence for each measure, with feedback being one source of evidence.
- The evaluation should allow flexibility for districts to take into consideration factors such as a principal's tenure state, years of experience, and years serving in a particular school.
- The evaluation tool should be a standardized rubric, in which all evaluation tools and data collection sources are research-based and psychometrically strong, with demonstrable validity.

Views expressed in this section reflect unique points to informing the decision making. Workgroup comments are folded in with Task Force comments. Not all required elements were discussed in depth.

Feedback from the Supervisor Plus Two Additional Stakeholders

Task Force members agreed the broad assessment must reflect feedback from the supervisor and at least two additional sources: teachers, students, parents/families, community stakeholders, and self-assessment. Formalized and standardized feedback from at least two additional sources would also be required. Additional sources were identified as teachers, students, parents/families, community stakeholders, and self-assessments.

At least one Task Force member felt differently about the use of surveys in assessing this domain of a principal's work: "This is quite unfair to New York City principals who really have no active direct supervisor. It would make no sense to agree to this under the current organization of NYCDOE as we have been positioned over the past few years to be *the only* buffer for teacher, parent, and community issues. Being pitted in such an adversarial role would lead to a vengeful use of noted 'surveys.'"

Standardized Rubric

All Task Force members agreed that principal behaviors, skills, and accomplishments should be measured using a standardized rubric developed by either districts or a third party. This would be based on the following measures:

Required Measures of Principal Performance

Discussion centered on key measures of principal performance, and the Task Force reached agreement that the four measures below should be required measures of a principal evaluation:

- Broad Assessment of Principal Leadership and Actions
- Assessment of Principal’s Skill in Supervising and Developing Teachers
- Progress Against Individual Professional Growth Goals
- Progress Against Ambitious and Measurable School Academic or Learning Environment Goals

Views expressed in this section reflect unique points to informing the decision making. Workgroup comments are folded in with Task Force comments.

Broad Assessment of Principal Leadership and Actions

There was strong agreement that this should be a required measure. Included in this measure is the measure of a principal’s skill in managing school operations and resources *to the extent that a principal has authority*. Task Force members recognized that the role of each principal varies and that, dependent on the amount of responsibility given to a principal, this dimension of the broader assessment would have to be taken into consideration. One Task Force member noted, “If the principal isn’t given [for example] a staffing budget, he can’t be held responsible for the operation costs of the school. If principals don’t have the authority to interview or hire teachers for their school, they shouldn’t be held responsible for the process. The same could be said for building maintenance issues.”

Assessment of Principal’s Skill in Supervising and Developing Teachers

All Task Force members strongly agreed that the principal evaluation must assess in some way the principal’s skill in supervising and developing teachers. Task Force members saw this as a major responsibility of principals, with all principals having some influence on teacher effectiveness. The group held that a principal evaluation “[should] include several possible measures of principal actions to support effective teaching, without requiring a specific one. [This] would keep leadership support of quality teaching a priority without being overly prescriptive.” The Task Force suggested that, depending on local context, a principal’s influence may be limited with regard to the principal’s ability to hire, fire, grant or deny tenures, compensate or retain teachers. Therefore, measures should be carefully selected to ensure fairness and efficacy, and districts should have the flexibility to select appropriate metrics, depending on principal authority. The Task Force identified several potential sources of evidence a district would have the flexibility to select:

- **Teacher Attendance**
- **Quality/Effectiveness of Teacher Evaluations**

- **Records showing that ineffective staff are dismissed after given a fair opportunity to improve**
- **Provision of opportunities for teachers to collaborate with colleagues**
- **Provision of opportunities for accomplished teachers to share their expertise and serve as coaches, mentors, etc.**
- **Skill in Improving Teacher Effectiveness over Time**—Many Task Force members agreed that this would be an appropriate metric. Where there was strong disagreement (with only one or two members agreeing) is whether teacher ratings of effectiveness or subcomponent scores could be aggregated into a principal evaluation measure.

Some Task Force members expressed concern that it would lead to double counting of student achievement or that it would be “psychometrically noisy.” One Task Force member noted, “Aggregating change scores is very problematic for evaluation purposes and is a poor reuse of the 40% student growth measure.”

- **Additional Suggestion**—A Task Force member suggested adding tenure accountability: “Make principals accountable for all of their tenure decisions, even after a teacher has left their school. Add a discrete timeframe, however; for example, the principal is accountable for that decision up to three years after a teacher has left his/her supervision.”

Progress Against Individual Professional Growth Goals

There was agreement by the Task Force that this measure should be required. Task Force members indicated that professional development goals would have to be grounded in clearly articulated leadership domains and standards that are research based and that promote positive student-achievement outcomes. The goals would be selected by the principal and approved by the supervisor. Both parties would determine how acquisition of skills and knowledge will be demonstrated as well as determine what evidence must be provided.

In support of this measure as a requirement, one Task Force member noted, “Professional development is an important element of the evaluation, and was emphasized in the Cohesive Leadership conversations earlier. The burden of professional development is on the evaluator as well as on the evaluated because there needs to be agreement between the two. It also puts the onus on superintendents to dedicate resources and time.” Another member commented, ““Ongoing growth and professional learning are critical for a principal and they should be required to do so, incentivized to do so, and evaluated for successful growth.”

Another Task Force member observed that growth goals can inspire and support innovation and risk taking on behalf of student achievement, and that they should be considered an important measure of a principal's performance.

The Task Force discussed whether it is appropriate to use this formative element for accountability purposes. One Task Force member commented, "There is no one with whom I could have this conversation as a New York City principal. My superintendent does not actually manage conversations of this nature and knows so little of my leadership strengths and weaknesses I don't know how we could make this component relevant." Another member voiced this concern: "While yearly professional goals are important for a principal and they need to be standards based, they should not be limited to student achievement outcomes. I am concerned that this language represents a form of double jeopardy for principals because they have already been accountable for the first 40%."

The Task Force also developed guidelines for principal professional growth and development, taken from the Cohesive Leadership System proposed guidelines.

Professional growth of the educational leader as demonstrated through actions and outcomes that impact student learning and achievement is an important objective of the educational leader evaluation/assessment process.

- Supervisors should provide specific and timely feedback through regularly scheduled meetings and ongoing communication.
- Evidence of the impact of professional growth efforts should be demonstrated
- The educational leader should engage in personal and collegial reflective practice that promotes professional growth.
- The educational leader's professional development and support should be differentiated based on need.
- Professional development for the educational leader should be identified through collaborative discussion with his/her supervisor.
- Professional growth should be tied to ISLLC Standards and built upon New York State's Professional Development Standards.

The Educational leader will develop and demonstrate a level of expertise by targeting a single ISLLC Standard and concentrating on one or more of its related functions.

- The targeted ISLLC Standard and its related functions should be selected by the educational leader and shared with his/her supervisor.
- The educational leader and his/her supervisor will determine how acquisition of knowledge and skills will be demonstrated.

- The educational leader should provide evidence demonstrating acquisition, enhancement and application of skills/expertise to his/her supervisor.

Progress Against Ambitious and Measurable School Academic or Learning Environment Goals

Task Force members agreed that this should be a required measure. Task Force members indicated that such goals would have to address substantive issues identified through data analysis and be jointly selected with the supervisor.

One Task Force member commented, “Setting goals with a supervisor can account for issues and resources, as well as determine where the principal is starting from.” Another member acknowledged, however, that “school-wide goals are often multi-year initiatives. Or circumstances may require a principal to put a school goal on hold for the year. A principal is quite often not the only person responsible for the attainment of school-wide goals.”

Two Task Force members stressed that in practice the use of goals as part of principal evaluation in NYC is not rigorous and most principals earn the full points available. Another member counters that the goals and objectives section of the PPR (worth 31 points out of 100) “demonstrates a full range of ratings by superintendents. Only 260 principals received full credit; 189 received ratings between 25 and 30; 447 scored between 20 and 24; 214 scored between 15 and 19; 59 scored between 10 and 14; and 117 scored below 10 points. These ratings for 2009-10 give evidence that the goals and objectives, developed collaboratively between principals and superintendent, can demonstrate rigor and accountability, and can provide an instrument to evaluate other components of a principal’s performance which progress reports, quality reviews, and a compliance list do not encompass.”

Sources of Evidence

There was broad consensus by the Task Force that a principal’s supervisor should take into consideration optional sources of evidence. The group agreed that there should be multiple sources of evidence for each measure and agreed to the following optional sources:

- School Visits
- Formalized and Standardized Feedback from Students, Parents, Teachers, Community Stakeholders
- Principal Self-Reflection
- Review of School Documents, Systems, and Records
- Results and Findings from Existing State Accountability Processes
- Other sources as determined locally

School Visits

While many Task Force members indicated that a school site visit is important to assessing a principal, others did not support this as a source of evidence. One member commented, “The evaluation should promote collaboration and communication. I have a concern regarding an outside evaluator who has no ongoing relationship with a principal.” A member noted, “I don’t believe the NYC Quality Review or any review like it has any value. Many times the principal and teachers prepare for the review in advance. You can not get a clear picture of the school or principal effectiveness this way.”

Formalized and Standardized Feedback from Students, Parents, Teachers, Community Stakeholders

This was an area where the Task Force was in agreement. The Task Force suggests using a model whereby the principal receives feedback from his or her supervisor and two additional sources. One Task Force member who supported the two additional sources noted, “I think a baseline of supervisor plus one other source is a good starting point as a minimal number of required sources of feedback. I would say that feedback from other sources is [already] routinely used as part of a principal’s evaluation.” Another Task Force member explained, “I would say at least three additional sources, because if left at two one could be self and one could be community stakeholders, for example. In that case, feedback from students and parents/families would be entirely missing. One could argue that these are the most important constituencies.” Yet another Task Force member commented, “Multiple measures are paramount. The supervisor must take into consideration various sources in addition to his/her perceptions.”

A few Task Force members disagreed with the number of additional sources. One stated, “I think supervisor plus one (not at least two other sources).” Another commented, “I don’t believe feedback should come from students.” Finally, another Task Force member observed, “If it takes three sources to assess a principal, the field will strongly feel that teachers should not be subjected to negative assessment based on only one source.”

Principal Self-Reflection

This source of evidence was identified as a possible option for evidence of professional growth.

Review of School Documents, Systems, and Records

The Task Force did not discuss this source of evidence in depth, but did give consideration to principal and supervisor negotiating and collaboratively deciding on the specific evidence.

Results and Findings from Existing State Accountability Processes

While ISLLC standards are the prevailing framework, attention to federal, state, and local improvement initiatives as sources of evidence of principal performance must be accommodated as well, as they reflect the reality of some principals' areas of responsibility. Existing accountability processes include, but are not limited to, the Special Education Performance Plan and NYSED's Differentiated Accountability School Quality Review. Including results and findings from accountability systems as potential sources of evidence encourages coherence and alignment.

Other Sources as Determined Locally

Districts should be granted flexibility to identify objective and measurable sources of evidence not specified in this list, in order to reflect the reality of the principalship in their districts.

Weighting of Evidence

The Task Force agreed that evidence should be scored holistically, with local districts determining the weight to be given to each element. The regulations should not mandate specific weights or numbers/ranges of points for particular measures within the 60%; instead, districts should be free to determine how the measures within the 60% are scored. This recommendation is made so that the evaluation system can accurately reflect the reality of the principalship in a given building or district, with the recognition that principals in different settings have authority over different domains and that a one-size-fits-all evaluation system would not suffice to develop school leaders.

Some Task Force members favored a holistic scoring system, in which the supervisor would have the discretion to consider each piece of feedback and other evidence in context. Some Task Force members felt strongly that, whatever scoring approach is used, supervisors should not be able to "throw out" feedback from other sources, or other evidence, and that allowing them to do so would undermine the validity of the evaluation process.

“Metrics”: Teacher of Record, Scoring, and Value-Added

Context

In February 2011, members of the Task Force began a focused discussion on policy recommendations regarding scoring and metrics items. The items associated with scoring and metrics that are being considered by the Task Force as the basis for making recommendations to the Board of Regents can be organized into four main areas: teacher of record policy; teacher and principal composite scoring bands; teacher and principal subcomponent scoring levels; and, teacher and principal state-wide growth measures.

General Principle

Purpose of Teacher and Principal Evaluation System

- To improve student results, all assumptions and consensus decisions should be tested empirically to determine whether they further that goal. The system must be one that is fair and equitable to those who will be evaluated.

Teacher of Record Policy

Teacher of Record policy covers decisions about which students count for which teachers' evaluations. SED needs to collect enhanced data from districts to provide New York State with flexibility in setting Teacher of Record policy.

The Task Force agreed with recommendations from SED that the policy to assign Teacher of Record would change as student management systems develop over the next few years, per the guidelines below:

- For those courses included in the 2010-11 school year collection, the Teacher of Record is the single teacher who is primarily and directly responsible for the student's year-long learning activities that are aligned to the performance measures of the course.
- For those courses included in the 2011-12 school year collection, Teachers of Record are those teachers who are primarily and directly responsible for the student's year-long learning activities that are aligned to the performance measures of the course. Note that in 2011-12, districts that have the capacity may provide more than one teacher of record for a course.
- The Task Force agreed that vendors should enhance their student management systems to allow policy flexibility in determining which students become part of which teachers' evaluations. Enhancements vendors will provide include the following teacher-student linkage elements:
 - course length;

- teacher-student linkage start/end dates;
 - student enrollment-teacher assignment linkage for course;
 - student attendance-teacher assignment linkage for course;
 - teacher-student instructional weightings; and,
 - a student exclude flag.
- A continuous enrollment weighting expressed as a percentage of available time that a student is enrolled in a course is more desirable than a cut-off point (for example 75%) below which a student does not “count” for a particular teacher. All students who are taught by a teacher should count to some degree in the teacher’s composite score, though such a system could not begin before 2012-13 at the earliest.
 - A student exclude flag can be reported for specified situations where a student should not be counted toward a teacher’s evaluation (subject to principal verification and approval).
 - Teachers must be involved in roster verification, though concern was expressed regarding the complexity involved with manual verification and system updates.

Some Task Force members were concerned with the feasibility of providing Teacher of Record in all subjects, but that this was particularly challenging for grades and subjects other than 4-8 ELA and math for 2011-12. Task Force members recognized the need for this data to support growth modeling, and suggested the need for state support to accomplish this in the pre-determined timeframe (e.g. a statewide verification process for teachers and schools).

A few Task Force members expressed concern with district capability to capture some linkage variables, in particular given that vendor systems will not be available until halfway through 2011-12 school year. It was noted that student management system vendors should be able to calculate percentage of available enrollment and attendance minutes so districts will not have to provide this information.

Subcomponent Scoring (State growth, local assessment, and other 60%)

The Task Force did not complete discussion on subcomponent scoring. They agreed on a number of principles described in this report, and they recognized that they are split on how prescriptive the state should be on the “local assessment” and “other 60%” subcomponents. There was intense discussion regarding whether the scoring of the 20% local and 60% other is subject to local collective bargaining. There is also a split regarding scoring for student growth measures.

The Task Force agreed that with regard to the subcomponent scores:

- They will require contextualization, either explicitly through regulations or implicitly through procedures. Uniform standards should guide use of a rubric for evaluation, with standards for whether and how contextual components, such as experience and conditions, should be taken into account.
- The process by which points are assigned in subcomponents must be transparent and provided in advance to those being rated.
- District plans should be made publicly available in electronic form and should specify how points will be assigned based on locally selected and other measures.
- The method for assigning subcomponent points should not be all or nothing, but rather the full range of available points should be possible under the scoring system.
- The Commissioner's regulations should at minimum, articulate factors to be considered by districts in assigning points to the subcomponents.
- There is a need for rigor and reliability in the application of the evaluation system—tied to training and certification of evaluators.

With regard to state assessments for measuring growth, some Task Force members believe that it is the Commissioner's responsibility both to determine appropriate assessments to be used and to assign points for teacher and principal evaluation in a uniform manner across the whole state. In the case of grades and subjects where there is no state assessment, some Task Force members believe that districts should have flexibility to choose a growth measure from a set of options. Others believe that it is the state's responsibility to determine the 20% student growth subcomponent of the evaluation and that this responsibility should not be delegated to school districts; there was concern that some districts would not have the capacity to make decisions at the local level that would be comparable to the points assigned to teachers with state assessments and that, therefore, the state needs to determine how growth in non-tested subjects will be measured and how the growth score will be calculated.

Discussion began regarding whether an alternative approach to the distribution of points could be used for teachers and principals in non-tested subjects. Some members thought options should be considered for non-tested subjects. One option suggested was that points for the state growth could be distributed in some way between the two other subcomponent categories, but discussion continued and consensus was not reached.

Task Force members considered examples of degrees of prescriptiveness for regulations around how the local and "other 60%" subcomponents should be scored

and agreed that at a minimum, regulations should articulate factors to be considered by districts in assigning points. Some Task Force members expressed concern with non-prescriptive regulations because this would not achieve the goal of a statewide, uniform system, while others felt the scoring approach for local assessment and particularly the 60% other subcomponents must be entirely locally determined. The Task Force had considerable discussion regarding the degree to which regulations should establish common scoring standards, in particular considering instances where districts are using the same locally selected assessments, common scoring rubrics, etc. A few Task Force members expressed concern with the comparability state-wide if there were no set standards. Other Task Force members expressed the need for flexibility and suggested districts should have the discretion to make decisions. For example, some members felt that if two teachers received the same evaluation on an observation rubric in two different districts the regulations should ensure the teachers receive the same (or similar) number of points towards their composite score, while other members felt that it was a local determination how the rubric results should be converted into points toward the composite score. The group recognized this is not an area where consensus will be reached.

Many Task Force members felt that the HEDI labels should not be used at the subcomponent level, only at the composite level and that if subcomponent levels are utilized they should have different terms (e.g. Levels 1-4). Most Task Force members felt that the state should provide text-based descriptions of four performance levels in all three subcomponents. The group was split about whether the state should also provide numerical guidance or establish scoring ranges for performance levels in subcomponents, disagreeing about the degree to which the subcomponent point assignment is up to local jurisdictions to determine or whether more state prescriptiveness is required to ensure equity and comparability. Task Force members agreed that the following chart of text-based descriptions is a starting point for further discussion. Many Task Force members agreed that the level-naming scheme should not be HEDI, but levels 1-4 is only an illustrative alternative.

Level	Growth	Locally-selected assessment results	Other (Teacher and Leader standards)
L1	Educator's students' results are well-below state average for similar students.	Educator's results do not achieve expectations for student learning.	Not approaching standard on most areas associated with district priorities within standards.
L2	Educator's students' results are below state average for similar students.	Educator's results partially achieve expectations for student learning.	Meets some standards. Needs improvement in areas associated with district priorities.
L3	Educator's students' results meet state average for similar students.	Educator's results achieve expectations for student learning.	Meets all or almost all standards associated with district priorities, may exceed some and may be in need of improvement in areas that are lower priority for the district.
L4	Educator's students' results are well-above state average for similar students.	Educator's results exceed expectations for student learning.	Meets all standards and exceeds many standards within district priorities.

Task Force members had much debate on the wording of the “local assessment growth or achievement” column in the chart above, particularly with regard to whether it should be labeled “district expectations” or “local expectations.” This debate centered on the degree to which the process for assignment of points is subject to collective bargaining, because some districts may value and weight the standards differently.

Composite Scoring Bands

The Task Force agreed that the bands for the composite score should be established in such a way that:

- the minimum and maximum scores that are used to determine the four effectiveness categories should be the same for all teachers and principals;
- a teacher or principal who scores substantially below expectations in each of the three components should receive an overall rating of ineffective and a person who scores substantially above expectations should receive a score of highly effective; and,
- a teacher or principal is not precluded from receiving a rating of effective solely because of growth results that are below expectations.

Task Force discussion suggested that the composite score should be calculated by the state after districts provide the scores for the locally-selected assessment and other

measures subcomponents. The Task Force considered numerous options for scoring bands at the composite level. After much deliberation, the group was able to narrow options for scoring bands as reflected in the table below for a 0-100 score range, but did not reach consensus on exact minimum and maximum scoring ranges to recommend for each rating category.

Rating Category	Band	Note
Ineffective (I)	0-X	X must be at least 39; some supported 49, and could go as high as 65.
Developing (D)	X-Y	X can be no lower than 40. Y is between 64 and 74.
Effective (E)	P-85	P is between 65 and 85.
Highly Effective (H)	86-100	Although some supported lower limit as high as 91.

Some Task Force members felt that an analogy to student grades could provide guidance for setting composite scoring bands, particularly if the range of scores is 0-100. They suggested that parents would feel that it makes sense to “grade” teachers and principals like students, especially if educators’ evaluations are supposed to represent their level of success with student learning. This analogy argues for highly effective band range from 90 or 91 to 100 like an “A” grade, and the ineffective category going from 0 to at least 50, if not 64. Others argued that this analogy is imperfect and too limiting and suggested that the score range be something other than 0-100, precisely to avoid the public thinking that the rating categories are essentially an A, B, C/D, F grading system. Some members felt that scoring ranges need to be set so that there can be differentiation of teachers within ranges; which they felt was not possible using tight ranges. The group did not reach consensus about whether or not to use a 100 point scale, or what the alternative would be. Suggestions included 0-160 since it bears no relation to tests; less than 100 since it might be easier to manage than large numbers of points; or some multiple of 10 (0-10.0; 0-200) because it may be easier to understand. There did appear to be consensus that SED’s communication about the system should place emphasis on the rating category rather than the numeric score (i.e. what is important is that a teacher has been rated “developing,” not that the teacher received a composite score of 54 or 222, if some alternate scale is used.).

Some Task Force members felt that the Commissioner should set the bands assuming the 60 “other points” or the 20 “local assessment” points will be “easy to get,” therefore arguing for the higher upper bound on the ineffective category and a higher lower bound on the highly effective category. Other Task Force members suggested that setting the bands in this manner would have the unintended effect of pushing districts to give higher points in these areas than they feel are really deserved. Other Task Force

members cautioned that setting a high cut score for effective (e.g. 65), suggests that teachers and principals will need to get more than half the points in all subcomponents to be considered effective. Some Task Force members cautioned that if the 60% other and the 20% locally-selected are set through a locally negotiated process it makes it extraordinarily difficult to make recommendations about the composite bands. Task Force members reiterated the fact that recommendations must stand up to empirical testing, further student achievement, and be fair to those who are evaluated.

Some Task Force members stated that setting the bands without more information or decisions about the three subcomponent scores is an exercise in guess-work and judgment, and not a precise science. Some Task Force members felt that the bands cannot be set without more detailed information about the state growth model, although this is not possible at this time. Many Task Force members felt that the initial setting of the bands may need to be reevaluated, and the Commissioner should review them over time and make recommendations as necessary. The Task Force members recognized that this won't be perfect at the beginning so there should be options for modifying if necessary.

Task Force members agree that the Commissioner should review the regulations at least bi-annually to determine whether cut scores should be modified. Some Task Force members do not believe annual modifications to the system are appropriate because there will be no stability to the system for teachers and principals to plan for their evaluations. Task Force members agree that the Commissioner should periodically report to the Regents the effectiveness of the system in evaluating teachers and principals.

Growth and Value-Added Modeling

The Task Force members agree that the methodology used to measure student growth on state assessments should be clear and transparent to educators and the public and should be provided to them prior to the start of the school year in which they are to be measured.

The Task Force members agree that the growth measures should:

- To the extent practical, incorporate statistical techniques that recognize test measurement error, and differences in the robustness of data under circumstances like small n sizes, particularly to ensure that teachers' and principals' effectiveness is not underestimated or overestimated and that teachers and principals have clear and transparent knowledge about how these

factors relate to their individual scores, and clear and transparent knowledge of the confidence intervals (or equivalent) for their individual results.

- Teachers should be involved in verifying their student rosters before evaluation scores are calculated.

General Principles

Student Growth Model

- A student growth methodology will be used to measure student growth on the grade 4-8 ELA and math assessments.
- A student growth score (SGS) will only be computed for students who have a minimum of two consecutive years of assessment results in a tested subject. This score will rank each student's performance compared to students with the same previous state assessment history in that subject.
- The methodology for establishing SGS's will use as much of the student's prior state academic results as SED determines is necessary to establish the student growth score.
- The methodology will adjust the SGS before assigning teacher and principal rating points so that a teacher or principal's student growth result takes into account student poverty and special needs status (SWD, ELL). This will be the Teacher or Principal student growth score (TSGS, PSGS)
- A "beta" model will be constructed for teachers and principals using data from before the 2011-12 school year for modeling purposes only. This will allow the Task Force, and the Board of Regents to understand how the model works before it is approved for "production use". If deemed desirable, beta results may be shared with individual educators for feedback and/or informational/training purposes.

Value-Added Model

- In future years, if approved by the Regents, SED will construct a teacher or principal VA score (TVA or PVA) using student growth scores and a wider range of student, classroom and/or school characteristics as policy and empirical analysis determine.

Teacher and Principal Student Growth (or VA) Scores

- The teacher or principal student growth scores will be calculated as means, medians, and standard deviations. The median is likely to be the primary number used in assigning effectiveness points in 2011-12. Where a teacher or principal is responsible for more than one class in a subject or grade, the

scores from different classes will be aggregated according to a methodology determined by SED to create an overall TSGS or PSGS.

- A minimum N size will be established below which a teacher or principal will not receive a TSGS or PSGS.
- The mean T or P SGS score will be converted into a teacher or principal effectiveness score of between 0 and 20, assuming a 100 point scale. The following describes, illustratively, four levels of performance, but this methodology will not be final until further data analysis is complete. NOTE: The following is an example only, and is not agreed upon by the Task Force.
 - A score of 0-5 means the teacher's or principal's students have demonstrated growth well below state average for similar students after application of a confidence interval.
 - A score of 6 -10 means the teacher's or principal's students have demonstrated growth below the state average for similar students after application of a confidence interval.
 - A score of 11 -15 means the teacher's or principal's students have demonstrated growth comparable to the statewide average for similar students after application of a confidence interval.
 - A score of 16 -20 means the teacher's or principal's students have demonstrated growth well above the state average for similar students after application of a confidence interval.
- Each teacher and principal will receive a detailed report about their TSGS and PSGS results in a format designed to be easy to understand and to provide information for developmental planning for each educator.
- Network teams will provide professional development in the SGS methodology and its implications for instructional decisions.

Task Force members agreed that almost all of the variables listed in the table below should be empirically tested by the modeling provider and recommendations made about what should be included in the state's value-added model. The variable that generated the most policy objection from people who felt it should not be included in the ultimate model was student ethnicity, with many Task Force members who responded to a survey indicating that it should not be included. The metrics group felt that empirical analysis of ethnicity should be conducted before any policy decisions are made.

State Data Elements

(items for which SED believes it has a valid data source now or in the near future)

Student Characteristics

- Student State assessment history

Other Characteristics

Classroom characteristics

- Disability indicators
 - Free/reduced price lunch or other poverty indicators
 - English language learner indicators
 - Ethnicity/race
 - Gender
 - % daily student attendance
 - Student suspension data
 - Retained in grade
 - Summer school attendance
 - Student new to school in a non-articulation year
 - Student age (especially to determine overage for grade)
 - Class size
 - % of students with each demographic characteristic in a class
- School characteristics**
- % of students with each demographic characteristic
 - Average class size
 - Grade configuration
 - Educator experience level in role

Some Task Force members asked for other items to be included in the value-added model that are not currently measurable consistently across the state, per the list below.

- More fine-grained categorization of students with disabilities and English Language Learners. (more than yes/no)
- School-based resources (including: technology, curriculum, extra-curricular options, guidance available)
- Health-related issues (including: mental, emotional, physical, nutrition)
- Leadership-related (including: experience of leadership, years as leader, control of leadership)
- Family-related (including: foster care, divorce)

Some Task Force members feel educators need to know exactly how the models will work and what scores they might have received in a previous year before the start of any year where the results will carry consequences. They also noted that setting scoring bands is difficult without more information about the results from the growth portion of evaluations. Many Task Force members feel that the timing of the growth model development means evaluation should not be implemented with consequences in 2011-12.

Some Task Force members expressed concern that a growth model that is based upon measuring changes in student performance compared to the performance of similar students is not truly a growth model. Some Task Force members expressed

reservations about having a growth model that would rank teachers and principals against each other in such a way that approximately half of the teachers and half of the principals each year would be considered below average in terms of student growth. They felt that all teachers and principals should have the opportunity to be effective, and should not be referenced against a norm.

Status Report on Growth/VA modeling

- The Task Force has established guiding principles for modeling, which are included in the previous pages of this report.
- The state has completed student growth percentile (SGP) calculations for all students based on 2009-10 school year data and earlier and median SGPs have been calculated for all schools in the state.
- Task Force researchers (Jonah Rockoff, Hamp Lankford and Jim Wyckoff) have agreed to provide illustrations of how the school-level SGP data could be converted into scores for principal evaluation later in spring 2011.
- Illustrative results of how SGP data can be converted into evaluation scores for teachers will be available in the middle of school year 2011-12. This timing results because teacher of record data will not be available to the state until summer 2011, and the provider of the statistical results for principals and teachers needs to be selected by an RFP that will be issued in spring 2011.

Evaluator Training

Discussion among Task Force members around the need for effective evaluator training for teacher and principal evaluators is ongoing. Central to the discussions has been the idea that training is paramount for evaluators, as well as evaluatees. The need for professional development around evaluation is extensive. The Board of Regents must consider granting flexibility in how training will be rolled out, but must also acknowledge the potential lack of human capacity to roll out training under tight timelines.

General consensus has been reached that:

1. All evaluators should be trained to observe teacher practice using the district's selected rubric.
2. Training should be designed to ensure inter-rater reliability.
3. Districts will be required to provide periodic retraining. Districts need to provide co-evaluators or independent raters to cover if there is not a trained evaluator for a school. There should be an ongoing process of communication. Training will show how an evaluator norms their evaluation.

To date, the following recommendations have been suggested based on consensus of Task Force members and underscored by the belief that sustainable training guidelines and regulations are critical to the success of principal and teacher evaluations.

Content of Evaluator Training

- NYSED should identify key evaluation skills and knowledge, distinct from the evaluation orientation specific to a proprietary evaluation tool, which all evaluators must be trained in. These would include but not be limited to goal-setting, communication skills, and teacher and leadership standards.
- All teacher and principal evaluators should be trained on evaluating performance relative to standards.
- All rubric vendors should include instructions that identify specific training and implementation details that evaluators need to be aware of when using them.
- All teachers and principals should receive training and orientation to the evaluation system used to measure their performance. This orientation should take place at the beginning of the school year.

- Training in evaluation should be ongoing and should include initial calibration and periodic re-calibration of all evaluators in order to ensure reliability.
- Technical assistance should be provided to all districts and evaluators to determine criteria for acceptable evidence for content validity.
- All districts and evaluators should receive training in student-growth and value-added models, holistic scoring models, and weighting options for evidence.

Instructional Delivery

- NYSED should develop guidance on training options. For example, the role and capacity of network teams should be defined, and other delivery options – including regional train-the-trainer capacity, consultants, professional organizations, leadership academies – should be defined.
- Evaluator training should reflect an understanding of adult learning styles and utilize a coaching model.

Timeline

- The timeline of regulations and guidance will dictate the timeline for training roll-out. It is currently assumed that network team training will occur in July or August 2011. A more fully expressed longer range timeline is warranted that expresses the frequency of training and/or updating network teams' skills and knowledge, especially to accommodate later roll-out of the evaluation systems.
- NYSED should provide a detailed timeline that identifies the following key components of training and orientation that will address the needs of the 2011-2012 school year:
 - Structured timeline for training network teams, district trainers, etc. in the near and far term
 - Identification of evaluation rubrics
 - Timeline for certification of evaluators
- The majority of the Task Force is concerned with the state's capacity to train evaluator and with the aggressive timeline in which this high-stakes training must take place. The Task Force recognized disparity in resources across districts to achieve high quality evaluations. Some Task Force members believe NYSED should allow a pilot period in order for districts to monitor the effectiveness of the training and the evaluation systems.

Regulations and Guidance

- Local flexibility should be granted to school districts to identify and provide training to all evaluators of teachers and principals.
- NYSED should differentiate between the obligations of network teams and school districts in terms of training and certification of evaluators.

Professional Development for Teachers and Principals

Discussion among Task Force members around the process for professional development and the annual professional performance review has yielded the following recommendations:

1. Professional growth goals should be tied to the performance assessments.
2. Professional growth should be tied to NYS teacher standards or ISLLC standards and build upon New York State's Professional Development Standards
3. Any teacher or principal rated as "developing" or "ineffective" through an annual professional performance review will have a teacher or principal improvement plan developed by the school district or board of cooperative educational services.
 - This improvement plan is to be formulated and commenced no later than ten days after the date on which teachers or principals are required to report prior to the opening of classes for the school year.
 - Such an improvement plan shall be consistent with the regulations of the commissioner and developed locally through negotiations per Civil Service Law.
 - The Task Force agreed that professional development should be a priority for *all* principals and teachers, regardless of their rating.
4. Professional development for *all* teachers and principals, as tied to professional growth goals and/or an improvement plan, should
 - be identified through collaborative discussion with the supervisor and include but not be limited to coaching, induction support, and differentiated professional development, which are to be locally established (per article 14 of Civil Service Law); and
 - include
 - identification of needed areas of improvement (based on the annual performance review);
 - a timeline for achieving improvement;
 - the manner in which improvement will be assessed;
 - pre-determined intervals that serve as benchmarks for measuring progress; and
 - where appropriate, differentiated activities and professional development to support the teacher's or principal's improvement in those areas.
5. The teacher and principal and his/her supervisor will determine how acquisition of knowledge and skills will be demonstrated.

6. Supervisors should provide specific and timely feedback to each teacher or principal through regularly scheduled meetings and ongoing communication.
7. The teacher or principal should engage in personal and collegial reflective practice that promotes professional growth.
8. Evidence of the impact of professional growth efforts should be demonstrated at the end of the year.
9. The teacher or principal should provide evidence demonstrating acquisition, enhancement and application of skills/expertise to his/her supervisor at the end of the year.

Appendices

Appendix A—List of Task Force Members and Workgroup Assignments

First Name	Last Name	Title	Office	Work Group(s)
Milton	Cofield	Co-Chair of Task Force and Vice Chancellor	NYS Board of Regents	n/a
Lester	Young	Co-Chair of Task Force	NYS Board of Regents	n/a
Heather	Adams	Assistant in Research and Educational Services	New York State United Teachers (NYSUT)	Metrics
Michele	Ashley	Principal	Jesse Owens School, PS26 in Brooklyn	Principals
Jackie	Bennett	Special Representative	United Federation of Teachers (UFT)	Metrics
Vicki	Bernstein	Executive Director of Teacher Recruitment & Quality	New York City Department of Education (NYC DOE)	Teachers
Jack	Bierwirth	Superintendent	Herricks Public Schools	Metrics
Crystal	Boling-Barton	Vice President	New York State Federation of School Administrators	Principals
Jane	Bullowa	Assistant Superintendent for Instructional Services	Ulster BOCES	Principals
Kirsten	Busch Johnson	Executive Director, Teacher Effectiveness	New York City Department of Education (NYC DOE)	Teachers
Joanna	Cannon	Executive Director of Talent Research and Data	New York City Department of Education (NYC DOE)	Metrics
Amy	Carpenter	Teacher	The Captain Manuel Rivera Jr. School, PS/MS 279	Locally Selected
Stacey	Caruso-Sharpe	Mathematics Teacher	Lynch Literacy Academy, Amsterdam City School District	Teachers
Nicole	Catapano	Coordinator for Data Analysis	Washington-Saratoga-Warren-Hamilton-Essex BOCES	Metrics
Florence	Chapin	Executive Director, Performance Management Strategies	New York City Department of Education (NYC DOE)	Principals
Cynthia	Cuthbert	Special Education Teacher	Monroe 2 - Orleans BOCES	Teachers
Jim	Dexter	District Superintendent	Washington-Saratoga-Warren-Hamilton-Essex BOCES	Metrics
Jane	Doherty	Teacher	WHEELS (Washington Heights Expeditionary Learning Outward Bound School)	Non-Tested

First Name	Last Name	Title	Office	Work Group(s)
Johanna	Duncan-Poitier	Chancellor's Deputy for the Education Pipeline and Vice Chancellor for Community Colleges	State University of New York (SUNY)	Principals
Mark	Fish	Principal	South Glens Falls CSD, Oliver W. Winch Middle School	Principals
Kerry	Flynn	Principal	Loudonville School	Non-Tested
Paul	Gasparini	Principal	Jamesville-Dewitt High School	Locally Selected Metrics
Aminda	Gentile	Vice President	United Federation of Teachers (UFT)	Locally Selected
Michael	Glover	District Superintendent	Genesee Valley Educational Partnership (BOCES)	Teachers
Patricia	Gould	Associate Counsel	New York State School Boards Association (NYSSBA)	Teachers
Janella	Hinds	Special Representative	United Federation of Teachers	Non-Tested, Metrics
Jere	Hochman	Superintendent	Bedford Central School District	Locally Selected
William	Hughes	Music Teacher	South Orangetown Central School District	Non-Tested
Jon	Hunter	Superintendent	Fairport Central School District	Non-Tested, Metrics
Dafny	Irizarry	ESL Teacher	Central Islip School District	Non-Tested
Lloyd	Jaeger	Superintendent	Millbrook Central School District	Principals
Demetra	Keane	Assistant Professor	Mercy College, School of Education	Teachers
Richard	Kessler	Executive Director	The Center for Arts Education	Non-Tested
Joan	Lucariello	University Dean of Education	The City University of New York (CUNY)	Teachers
John	Mahony	Career and Technical Teachers	Questar III BOCES, Aviation Technology	Non-Tested
Robert	McClure	Superintendent-in-Residence, Professional Development Team	New York State Council of School Superintendents (NYSCOSS)	Principals
Michael	McDermott	Principal	Scarsdale Middle School	Principals
Peter	McNally	Executive Vice President	Council of Supervisors and Administrators	Principals
Michael	Mendel	Secretary	United Federation of Teachers (UFT)	Teachers
Paul	Mooney	Principal	Jefferson-Lewis BOCES, Howard G. Sackett Technical Center	Non-Tested
Maria	Neira	Vice President	New York State United Teachers	Teachers, Metrics
Donald	Nickson	Deputy Executive Director	School Administrators Ass'n of New York State	Locally Selected

First Name	Last Name	Title	Office	Work Group(s)
Sonia	Nieves	CSA Supervisory Support Program/Principal	Council of Supervisors and Administrators/DOE	Principals
Christopher	Ogno	Principal	PS 247, The New York City Partnership School	Locally Selected, Metrics
Terry	Orr	Faculty/President	Bank Street College of Education/Metropolitan Council of Educational Admin. Programs	Principals
Ruth	Pagerey	Asst. Dean for Professional Studies	State University of New York (SUNY)	Principals
Joseph	Perez	Teacher/Member	Southern Westchester BOCES/Professional Standards and Practices Board	Teachers
Viviana	Perez	Social Studies Teacher	WHEELS (Washington Heights Expeditionary Learning Outward Bound School)	Non-Tested
John	Picarello	Principal	Nassau BOCES	Teachers
Joel	Ratner	District Music Coordinator	Brentwood Public Schools	Non-Tested
Deborah	Rider	Principal	Rochester City School District	Principals
Lori	Rider	Teacher	Avon Central School District	Teachers
Karen	Rock	ELA Instructor	Plattsburgh City School District	Locally Selected
Jeff	Rozran	Teacher/President	Syosset Central School District/Syosset Teachers' Association	Non-Tested
Anne	Scott	Chief of Schools	Rochester City School District	Teachers
Karen	Siegel	Principal	South Huntington School District	Teachers
Pilar	Sokol	Deputy General Counsel	New York State School Boards Association	Teachers, Metrics
Penny	VanDusen	Elementary Education Teacher	Utica City School District	Locally Selected
Elizabeth	Van Nest	Vice President for Legal Affairs and General Counsel	Commission on Independent Colleges and Universities (clcu)	Locally Selected
David	Vazquez	Principal	The Urban Assembly Bronx Studio School for Writers and Artists	Principals
Thomas	Vitale	Principal	Buffalo Public Schools, Occupational Training Center	Principals
Lawrence	Waite	Manager of Educational Services	New York State United Teachers (NYSUT)	Teachers
James	Waslawski	Principal	Captain Manuel Rivera Jr. School, PS/MS 279	Teachers
Sue	Winer	District Coach for Collaborative Coaching and Learning	Syracuse City Schools	Teachers, Metrics
Jay	Worona	General Counsel	New York State School Boards Association	Principals, Metrics

Appendix B—Advisors, Presenters, Staff, and Fellows

Research Advisors

Hamilton Lankford, SUNY Albany
Jonah Rockoff, Columbia University
James Wyckoff, Curry School of Education, University of Virginia
Damian Betebenner, Center for Assessment
Scott Marion, Center for Assessment

Expert Presenters

Robin Gelinias, Education Counsel
Laura Goe, Educational Testing Service
Matt Keleman, New Leaders for New Schools
Rob Meyer, Value-Added Research Center, University of Wisconsin
The New Teacher Project

NYSED Staff

David Steiner, Commissioner
John King, Senior Deputy Commissioner

David Abrams	Amy Lamitie
Allison Armour-Garb	Lisa Luderman
Jim Baldwin	Meg McNiff
Robert Bentley	Gail Moon
Elena Bruno	Peg Rivers
Anne DeFiglio	Shannon Roberson
Beth Dievendorf	Ira Schwartz
Barbara Downs	Ken Slentz
Joseph Frey	Jean Stevens
Concetta Gallo	Ken Wagner
Debbie Horton	Monica Young
Vicki Jones	

Regents Research Fellows

Kristen Huff
Amy McIntosh
Julia Rafal
Peter Swerdzewski

Appendix C—Work Cited and Works Reviewed

- Baker, Eva L., Barton, Paul E., Darling-Hammond, Linda, Haertel, Edward, Ladd, Helen F., Linn, Robert L., Ravitch, Diane, Rothstein, Richard, Shavelson, Richard J., and Shepard, Lorrie A. Problems with the Use of Student Test Scores to Evaluate Teachers. Issue brief #278. Washington, DC: Economic Policy Institute, 2010. http://epi.3cdn.net/b9667271ee6c154195_t9m6ijj8k.pdf.
- Building a Breakthrough Framework for Educator Evaluation in the Commonwealth. Task Force Report. Massachusetts Department of Elementary and Secondary Education. March 2011. www.doe.mass.edu
- Beers, Donald E., Delaware Department of Education: Delaware Performance Appraisal System. Second Edition (DPASII) Chicago: Progress Education Corporation, 2009. www.progresseducation.com
- Brown-Sims, Melissa. Evaluating Principals, National Comprehensive Center for Teacher Quality. July 2010.
http://www.tqsource.org/publications/KeyIssue_PrincipalAssessments.pdf.
- Corcoran, Sean. Can Teachers Be Evaluated by Their Students' Test Scores? Should They Be? The Use of Value-Added Measure of Teacher Effectiveness in Policy and Practice. Providence: Annenberg Institute for School Reform at Brown University, 2010.
- Danielson, Charlotte. "Assessing Teacher Effectiveness." February 2011. Powerpoint Presentation.
- District of Columbia Public Schools. IMPACT: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel.
www.dcps.dc.gov
- Educational Leadership Policy Standards: 2008 As Adopted by the National Policy Board for Educational Administration. Policy Paper. Council of Chief State School Officers.
http://www.ccsso.org/Resources/Publications/Educational_Leadership_Policy_Standards_ISLLC_2008_as_Adopted_by_the_National_Policy_Board_for_Educational_Administration.html.
- Enhancing Teacher Evaluation: Effective Practices for Evaluating Teachers of All Students. National Comprehensive Center for Teacher Quality, April 14, 2010.
<http://www.tqsource.org/webcasts/evaluatingTeachers/webcast/>.
- Erenben, Cara. "Should Student Test Scores Be Used to Evaluate Teachers?" Eschoolnews.com. January 12, 2011.

<http://www.eschoolnews.com/2011/01/12/should-student-test-scores-be-used-to-evaluate-teachers/>.

Evaluating Principals: Balancing Accountability with Professional Growth. New Leaders for New Schools. White Paper. <http://www.nlins.org/evaluating-principals.jsp>.

Glazerman, Steven, Goldhaber, Dan, Loeb, Susanna, Raudenbush, Stephen, Steiger Douglas, and Whitehurst, Grover. Evaluating Teachers: The Important Role of Value Added. The Brown Center on Education Policy at Brookings, November 17, 2010.

http://www.brookings.edu/reports/2010/1117_evaluating_teachers.aspx.

Goe, Laura. "Multiple Measures of Teacher Effectiveness." National Comprehensive Center for Teacher Quality, December 3, 2010.

<http://www.tqsource.org/publications/practicalGuide.pdf>.

Goldschmidt, Pete. "Practical Considerations for Using Value Added Models for Monitoring Teacher Effectiveness." Sacramento, CA. October 2008. Powerpoint Presentation at Value Added Modeling Meeting sponsored by the National Center for Research on Evaluation, Standards, & Student Testing.

Harris, Douglas N. Value Added Measures of Education Performance: Clearing Away the Smoke and Mirrors. Policy Brief. Stanford, CA: PACE School of Education, Stanford University, 2010.

Interim Report. New Jersey Educator Effectiveness Task Force. March 1, 2011.

Iowa School Leaders with Support from School Administrators of Iowa and The Wallace Foundation. Principal Leadership Performance Review: A Systems Approach.

Kelemen, Matt. "Student Outcomes for Principal Evaluation- Presentation to the NY Regents Task Force on Teacher and Principal Effectiveness." Power Point presentation sponsored by New Leaders for New Schools, January 20, 2011.

<http://www.nlins.org/evaluating-principals.jsp>.

Li, Ying, Marion, Scott, Perie, Marianne, and Gong, Brian. "An Approach for Evaluating the Technical Quality of Interim Assessments." Peabody Journal of Education 85.2 (2010): 163-85. <http://dx.doi.org/10.1080/01619561003685304>.

Little, Olivia, Goe, Laura, and Bell, Courtney. A Practical Guide to Evaluating Teacher Effectiveness. National Comprehensive Center for Teacher Quality, April 2009.

<http://www.tqsource.org/publications/practicalGuide.pdf>.

McREL's Principal Evaluation System. Mid Continent Research for Education and Learning. www.mcrel.org.

- Measuring Principal Performance: How Rigorous Are Commonly Used Principal Performance Assessment Instruments? Issue brief. Learning Point Associates, February 2010. www.learningpt.org/expertise/educatorquality/.
- Measuring School Performance and the Efficacy of Instructional Practices: Value Added Essentials. Dr. Robert Meyer. Value Added Research Center. Webcast http://varc.wceruw.org/tutorials/Introduction_to_VA/START.html.
- Meyer, Robert H. and Dokumaci, Emin. Value-Added Models and the Next Generation of Assessments. Educational Testing Service- Center for K-12 Assessment and Performance Management, Dec. 2009. www.k12center.org/publications.html.
- Milanowski, Tony, and Patrick Schuermann. "Principal Evaluation." Bethesda, MD. Powerpoint Presentation at the Teacher Incentive Fund Grantee Meeting. June 8-9, 2009.
- Milanowski, Tony, and Patrick Schuermann. "Teacher and Principal Observations: Concurrent Session 3-A." Bethesda, MD. Powerpoint Presentation at the Teacher Incentive Fund Grantee Meeting. June 5-6, 2008.
- Perie, Marianne, Gong Brian, and Marion, Scott. "Moving Toward a Comprehensive Assessment System- A Framework for Considering Interim Assessments." Educational Measurement: Issues and Practice 28.3 (2009): 5-13.
- Peterson, William and Rothstein, Richard. Let's Do The Numbers- Department of Educations "Race to the Top" Program Offers Only a Muddled Path to the Finish Line. Issue brief. Washington, DC: Economic Policy Institute, 2010.
- Porter, Andrew C., Elliott, Stephen N., Goldring, Ellen, May, Henry, Murphy, Joseph, and Polikoff, Morgan S.. Vanderbilt Assessment of Leadership in Education Technical Manual. Vanderbilt University, November 25, 2008. www.valed.com.
- Prince, Cynthia D. "The Other 69 Percent: Designing a Performance-Based Pay System for Teachers of Non-Tested Subjects and Grades, English Language Learners, and Students with Disabilities." Power Point presentation at the 2008 TIF Grantee Meeting. Vanderbilt University. June 5-6, 2008.
- Principal Effectiveness: A New Principalship to Drive Student Achievement, Teacher Effectiveness and School Turnarounds. New Leaders for New Schools, 2009. www.nlms.org/uef.jsp.
- Principal Leadership Actions: A Guide to Assessing Principal Actions That Drive Dramatic Gains in Student Achievement. New Leaders for New Schools. <http://www.nlms.org/uef.jsp>.
- Rochester City School District. CIA/S ASAR Handbook for the Supervision and Evaluation of Administrators. 2010-2011.

- Rothstein, Richard. "Fact-Challenged Policy." Web log post. National Journal Experts Blog. March 7, 2011. <http://education.nationaljournal.com/2011/03/focusing-on-teacher-effectiven.php#1903160>.
- Teacher Evaluation 2.0. The New Teacher Project, October 2010. <http://tntp.org/files/Teacher-Evaluation-Oct10F.pdf>.
- Weishug, Daniel, Keeling, David, Mulhern, Jennifer, and Sexton, Jennifer. The Widget Project: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness. The New Teacher Project, 2009. http://tntp.org/files/TheWidgetEffect_2nd_ed.pdf
- Winerip, Michael. "Evaluating New York Teachers, Perhaps the Numbers Do Lie." New York Times. March 6, 2011. www.nytimes.com/2011/03/07/education/07winerio.html?emc=tnt&tntemail1=y.

Appendix D—Chapter 103 of the Laws of 2010

LAWS OF NEW YORK, 2010

CHAPTER 103

1 AN ACT to amend the education law, in relation to the evaluation of
2 teachers and principals; and to amend the education law, in relation
3 to authorizing school districts to contract with educational partner-
4 ship organizations to turn around certain low-performing schools

5

6 Became a law May 28, 2010, with the approval of the Governor.

7 Passed by a two-thirds vote.

8

9 The People of the State of New York, represented in Senate and Assem-
10 bly, do enact as follows:

11

12 Section 1. The education law is amended by adding a new section 3012-c
13 to read as follows:

14 § 3012-c. Annual professional performance review of classroom teachers
15 and building principals. 1. Notwithstanding any other provision of law,
16 rule or regulation to the contrary, the annual professional performance
17 reviews of all classroom teachers and building principals employed by
18 school districts or boards of cooperative educational services shall be
19 conducted in accordance with the provisions of this section. Such
20 performance reviews which are conducted on or after July first, two
21 thousand eleven, or on or after the date specified in paragraph c of
22 subdivision two of this section where applicable, shall include measures
23 of student achievement and be conducted in accordance with this section.
24 Such annual professional performance reviews shall be a significant
25 factor for employment decisions including but not limited to, promotion,
26 retention, tenure determination, termination, and supplemental compen-
27 sation, which decisions are to be made in accordance with locally devel-
28 oped procedures negotiated pursuant to the requirements of article four-
29 teen of the civil service law. Such performance reviews shall also be a
30 significant factor in teacher and principal development, including but
31 not limited to, coaching, induction support and differentiated profes-
32 sional development, which are to be locally established in accordance
33 with procedures negotiated pursuant to the requirements of article four-
34 teen of the civil service law.

35 2. a. The annual professional performance reviews conducted pursuant
36 to this section for classroom teachers and building principals shall
37 differentiate teacher and principal effectiveness using the following
38 quality rating categories: highly effective, effective, developing and
39 ineffective, with explicit minimum and maximum scoring ranges for each
40 category, as prescribed in the regulations of the commissioner. Such
41 annual professional performance reviews shall result in a single compos-
42 ite teacher or principal effectiveness score, which incorporates multi-
43 ple measures of effectiveness related to the criteria included in the
44 regulations of the commissioner. Except for the student growth measures
45 prescribed in paragraphs e, f and g of this subdivision, the elements
46 comprising the composite effectiveness score shall be locally developed,
47 consistent with the standards prescribed in the regulations of the
48 commissioner, through negotiations conducted, pursuant to the require-
49 ments of article fourteen of the civil service law.

50

51 EXPLANATION—Matter in italics is new; matter in brackets [] is old law
52 to be omitted.

1 CHAP. 103 2

2

3 b. Annual professional performance reviews conducted by school
4 districts on or after July first, two thousand eleven of classroom
5 teachers of common branch subjects or English language arts or mathemat-
6 ics in grades four to eight and all building principals of schools in
7 which such teachers are employed shall be conducted pursuant to this
8 subdivision and shall use two thousand ten—two thousand eleven school
9 year student data as the baseline for the initial computation of the
10 composite teacher or principal effectiveness score for such classroom
11 teachers and principals.

12 c. Annual professional performance reviews conducted by school
13 districts or boards of cooperative educational services on or after July
14 first, two thousand twelve of all classroom teachers and all building
15 principals shall be conducted pursuant to this subdivision and shall use

16 two thousand eleven—two thousand twelve school year student data as the
17 baseline for the initial computation of the composite teacher or princi-
18 pal effectiveness score for such classroom teachers and principals. For
19 purposes of this section, an administrator in charge of an instructional
20 program of a board of cooperative educational services shall be deemed
21 to be a building principal.

22 d. Prior to any evaluation being conducted in accordance with this
23 section, each individual who is responsible for conducting an evaluation
24 of a teacher or building principal shall receive appropriate training in
25 accordance with the regulations of the commissioner of education.

26 e. For annual professional performance reviews conducted in accordance
27 with paragraph b of this subdivision in the two thousand eleven—two
28 thousand twelve school year, forty percent of the composite score of
29 effectiveness shall be based on student achievement measures as follows:
30 (i) twenty percent of the evaluation shall be based upon student growth
31 data on state assessments as prescribed by the commissioner or a compa-
32 rable measure of student growth if such growth data is not available;
33 and (ii) twenty percent shall be based on other locally selected meas-
34 ures of student achievement that are determined to be rigorous and
35 comparable across classrooms in accordance with the regulations of the
36 commissioner and as are developed locally in a manner consistent with
37 procedures negotiated pursuant to the requirements of article fourteen
38 of the civil service law.

39 f. For annual professional performance reviews conducted in accordance
40 with paragraph c of this subdivision in any school year prior to the
41 first school year for which the board of regents has approved use of a
42 value-added growth model, but not earlier than the two thousand twelve-
43 -two thousand thirteen school year, forty percent of the composite score
44 of effectiveness shall be based on student achievement measures as
45 follows: (i) twenty percent of the evaluation shall be based upon
46 student growth data on state assessments as prescribed by the commis-
47 sioner or a comparable measure of student growth if such growth data is
48 not available; and (ii) twenty percent shall be based on other locally
49 selected measures of student achievement that are determined to be
50 rigorous and comparable across classrooms in accordance with the regu-
51 lations of the commissioner and as are developed locally in a manner
52 consistent with procedures negotiated pursuant to the requirements of
53 article fourteen of the civil service law.

54 g. For annual professional performance reviews conducted in accordance
55 with paragraph c of this subdivision in the first school year for which
56 the board of regents has approved use of a value-added growth model and
57 thereafter, forty percent of the composite score of effectiveness shall
58 be based on student achievement measures as follows: (i) twenty-five

1 3 CHAP. 103

2

3 percent of the evaluation shall be based upon student growth data on
4 state assessments as prescribed by the commissioner or a comparable
5 measure of student growth if such growth data is not available; and (ii)
6 fifteen percent shall be based on other locally selected measures of
7 student achievement that are determined to be rigorous and comparable
8 across classrooms in accordance with the regulations of the commissioner
9 and as are locally developed in a manner consistent with procedures
10 negotiated pursuant to the requirements of article fourteen of the civil
11 service law. The department shall develop the value-added growth model
12 and shall consult with the advisory committee established pursuant to
13 subdivision seven of this section prior to recommending that the board
14 of regents approve its use in evaluations.

15 h. The remaining percent of the evaluations, ratings and effectiveness
16 scores shall be locally developed, consistent with the standards
17 prescribed in the regulations of the commissioner, through negotiations
18 conducted pursuant to article fourteen of the civil service law.

19 i. For purposes of this section, student growth means the change in
20 student achievement for an individual student between two or more points
21 in time.

22 3. Nothing in this section shall be construed to excuse school
23 districts or boards of cooperative educational services from complying
24 with the standards set forth in the regulations of the commissioner for
25 conducting annual professional performance reviews of classroom teachers
26 or principals, including but not limited to required quality rating
27 categories, in conducting evaluations prior to July first, two thousand
28 eleven, or, for classroom teachers or principals subject to paragraph c
29 of subdivision two of this section, prior to July first, two thousand

30 twelve.

31 4. Notwithstanding any other law, rule or regulation to the contrary,
32 upon rating a teacher or a principal as developing or ineffective
33 through an annual professional performance review conducted pursuant to
34 subdivision two of this section, the school district or board of cooper-
35 ative educational services shall formulate and commence implementation
36 of a teacher or principal improvement plan for such teacher or principal
37 as soon as practicable but in no case later than ten days after the date
38 on which teachers are required to report prior to the opening of classes
39 for the school year. Such improvement plan shall be consistent with the
40 regulations of the commissioner and developed locally through negoti-
41 ations conducted pursuant to article fourteen of the civil service law.
42 Such improvement plan shall include, but need not be limited to, iden-
43 tification of needed areas of improvement, a timeline for achieving
44 improvement, the manner in which improvement will be assessed, and,
45 where appropriate, differentiated activities to support a teacher's or
46 principal's improvement in those areas.

47 5. An appeals procedure shall be locally established in each school
48 district and in each board of cooperative educational services by which
49 the evaluated teacher or principal may only challenge the substance of
50 the annual professional performance review, the school district's or
51 board of cooperative educational services' adherence to the standards
52 and methodologies required for such reviews, pursuant to this section,
53 the adherence to the regulations of the commissioner and compliance with
54 any applicable locally negotiated procedures, as well as the school
55 district's or board of cooperative educational services' issuance and/or
56 implementation of the terms of the teacher or principal improvement
57 plan, as required under this section. The specifics of the appeal proce-
58 dure shall be locally established through negotiations conducted pursu-

1 CHAP. 103 4

2

3 ant to article fourteen of the civil service law. An evaluation which is
4 the subject of an appeal shall not be sought to be offered in evidence
5 or placed in evidence in any proceeding conducted pursuant to either
6 section three thousand twenty-a of this article or any locally negoti-
7 ated alternate disciplinary procedure, until the appeal process is
8 concluded.

9 6. For purposes of disciplinary proceedings pursuant to sections three
10 thousand twenty and three thousand twenty-a of this article, a pattern
11 of ineffective teaching or performance shall be defined to mean two
12 consecutive annual ineffective ratings received by a classroom teacher
13 or building principal pursuant to annual professional performance
14 reviews conducted in accordance with the provisions of this section.

15 7. The regulations adopted pursuant to this section shall be developed
16 in consultation with an advisory committee consisting of representatives
17 of teachers, principals, superintendents of schools, school boards,
18 school district and board of cooperative educational services officials
19 and other interested parties. The regulations shall also take into
20 account any (i) professional teaching standards; (ii) standards for
21 professional contexts; and (iii) standards for a continuum of system
22 support for teachers and principals developed in consultation with the
23 advisory committee. Regulations promulgated pursuant to this section
24 shall be effective no later than July first, two thousand eleven, for
25 implementation in the two thousand eleven-two thousand twelve school
26 year.

27 8. Notwithstanding any other provision of law, rule or regulation to
28 the contrary, all collective bargaining agreements applicable to class-
29 room teachers or building principals entered into after July first, two
30 thousand ten shall be consistent with requirements of this section.
31 Nothing in this section shall be construed to abrogate any conflicting
32 provisions of any collective bargaining agreement in effect on July
33 first, two thousand ten during the term of such agreement and until the
34 entry into a successor collective bargaining agreement, provided that
35 notwithstanding any other provision of law to the contrary, upon expira-
36 tion of such term and the entry into a successor collective bargaining
37 agreement the provisions of this section shall apply. Furthermore, noth-
38 ing in this section or in any rule or regulation promulgated hereunder
39 shall in any way, alter, impair or diminish the rights of a local
40 collective bargaining representative to negotiate evaluation procedures
41 in accordance with article fourteen of the civil service law with the
42 school district or board of cooperative educational services.

43 § 2. Subdivisions 1 and 3 and paragraph a of subdivision 4 of section

44 3020 of the education law, subdivision 1 as added by chapter 691 of the
45 laws of 1994, subdivision 3 as added by chapter 3 of the laws of 2000
46 and paragraph a of subdivision 4 as added by section 1 of part J of
47 chapter 93 of the laws of 2002, are amended to read as follows:
48 1. No person enjoying the benefits of tenure shall be disciplined or
49 removed during a term of employment except for just cause and in accord-
50 ance with the procedures specified in section three thousand twenty-a of
51 this article or in accordance with alternate disciplinary procedures
52 contained in a collective bargaining agreement covering his or her terms
53 and conditions of employment that was effective on or before September
54 first, nineteen hundred ninety-four and has been unaltered by renegoti-
55 ation, or in accordance with alternative disciplinary procedures
56 contained in a collective bargaining agreement covering his or her terms
57 and conditions of employment that becomes effective on or after Septem-
58 ber first, nineteen hundred ninety-four; provided, however, that any

1 5 CHAP. 103

2

3 such alternate disciplinary procedures contained in a collective
4 bargaining agreement that becomes effective on or after September first,
5 nineteen hundred ninety-four, must provide for the written election by
6 the employee of either the procedures specified in such section three
7 thousand twenty-a or the alternative disciplinary procedures contained
8 in the collective bargaining agreement and must result in a disposition
9 of the disciplinary charge within the amount of time allowed therefor
10 under such section three thousand twenty-a; and provided further that
11 any alternate disciplinary procedures contained in a collective bargain-
12 ing agreement that becomes effective on or after July first, two thou-
13 sand ten shall provide for an expedited hearing process before a single
14 hearing officer in accordance with subparagraph (i-a) of paragraph c of
15 subdivision three of section three thousand twenty-a of this article in
16 cases in which charges of incompetence are brought based solely upon an
17 allegation of a pattern of ineffective teaching or performance as
18 defined in section three thousand twelve-c of this article and shall
19 provide that such a pattern of ineffective teaching or performance shall
20 constitute very significant evidence of incompetence which may form the
21 basis for just cause removal.

22 3. Notwithstanding any inconsistent provision of law, the procedures
23 set forth in section three thousand twenty-a of this article and subdi-
24 vision seven of section twenty-five hundred ninety-j of this chapter may
25 be modified or replaced by agreements negotiated between the city school
26 district of the city of New York and any employee organization repres-
27 enting employees or titles that are or were covered by any memorandum of
28 agreement executed by such city school district and the council of
29 supervisors and administrators of the city of New York on or after
30 December first, nineteen hundred ninety-nine. Where such procedures are
31 so modified or replaced: (i) compliance with such modification or
32 replacement procedures shall satisfy any provision in this chapter that
33 requires compliance with section three thousand twenty-a, (ii) any
34 employee against whom charges have been preferred prior to the effective
35 date of such modification or replacement shall continue to be subject to
36 the provisions of such section as in effect on the date such charges
37 were preferred, (iii) the provisions of subdivisions one and two of this
38 section shall not apply to agreements negotiated pursuant to this subdi-
39 vision, and (iv) in accordance with paragraph (e) of subdivision one of
40 section two hundred nine-a of the civil service law, such modification
41 or replacement procedures contained in an agreement negotiated pursuant
42 to this subdivision shall continue as terms of such agreement after its
43 expiration until a new agreement is negotiated; provided that any alter-
44 nate disciplinary procedures contained in a collective bargaining agree-
45 ment that becomes effective on or after July first, two thousand ten
46 shall provide for an expedited hearing process before a single hearing
47 officer in accordance with subparagraph (i-a) of paragraph c of subdivi-
48 sion three of section three thousand twenty-a of this article in cases
49 in which charges of incompetence are brought against a building princi-
50 pal based solely upon an allegation of a pattern of ineffective teaching
51 or performance as defined in section three thousand twelve-c of this
52 article and shall provide that such a pattern of ineffective teaching or
53 performance shall constitute very significant evidence of incompetence
54 which may form the basis for just cause removal of the building princi-
55 pal. Notwithstanding any inconsistent provision of law, the commission-
56 er [of education] shall review any appeals authorized by such modifica-
57 tion or replacement procedures within fifteen days from receipt by such

58 commissioner of the record of prior proceedings in the matter subject to

1 CHAP. 103 6

2

3 appeal. Such review shall have preference over all other appeals or
4 proceedings pending before such commissioner.

5 a. Notwithstanding any inconsistent provision of law, the procedures
6 set forth in section three thousand twenty-a of this article and subdi-
7 vision seven of section twenty-five hundred ninety-j of this chapter may
8 be modified by agreements negotiated between the city school district of
9 the city of New York and any employee organization representing employ-
10 ees or titles that are or were covered by any memorandum of agreement
11 executed by such city school district and the united federation of
12 teachers on or after June tenth, two thousand two. Where such proce-
13 dures are so modified: (i) compliance with such modified procedures
14 shall satisfy any provision of this chapter that requires compliance
15 with section three thousand twenty-a of this article; (ii) any employee
16 against whom charges have been preferred prior to the effective date of
17 such modification shall continue to be subject to the provisions of such
18 section as in effect on the date such charges were preferred; (iii) the
19 provisions of subdivisions one and two of this section shall not apply
20 to agreements negotiated pursuant to this subdivision, except that no
21 person enjoying the benefits of tenure shall be disciplined or removed
22 during a term of employment except for just cause; and (iv) in accord-
23 ance with paragraph (e) of subdivision one of section two hundred nine-a
24 of the civil service law, such modified procedures contained in an
25 agreement negotiated pursuant to this subdivision shall continue as
26 terms of such agreement after its expiration until a new agreement is
27 negotiated; and provided further that any alternate disciplinary proce-
28 dures contained in a collective bargaining agreement that becomes effec-
29 tive on or after July first, two thousand ten shall provide for an expe-
30 dited hearing process before a single hearing officer in accordance with
31 subparagraph (i-a) of paragraph c of subdivision three of section three
32 thousand twenty-a of this article in cases in which charges of incompe-
33 tence are brought based solely upon an allegation of a pattern of inef-
34 fective teaching or performance as defined in section three thousand
35 twelve-c of this article and shall provide that such a pattern of inef-
36 fective teaching or performance shall constitute very significant
37 evidence of incompetence which may form the basis for just cause
38 removal.

39 § 3. Paragraph (c) of subdivision 2 of section 3020-a of the education
40 law, as amended by chapter 691 of the laws of 1994, is amended to read
41 as follows:

42 (c) Within ten days of receipt of the statement of charges, the
43 employee shall notify the clerk or secretary of the employing board in
44 writing whether he or she desires a hearing on the charges and when the
45 charges concern pedagogical incompetence or issues involving pedagogical
46 judgment, his or her choice of either a single hearing officer or a
47 three member panel, provided that a three member panel shall not be
48 available where the charges concern pedagogical incompetence based sole-
49 ly upon a teacher's or principal's pattern of ineffective teaching or
50 performance as defined in section three thousand twelve-c of this arti-
51 cle. All other charges shall be heard by a single hearing officer.

52 § 4. Paragraph a of subdivision 3 of section 3020-a of the education
53 law, as amended by chapter 691 of the laws of 1994, is amended to read
54 as follows:

55 a. Notice of hearing. Upon receipt of a request for a hearing in
56 accordance with subdivision two of this section, the commissioner [of]
57 [education] shall forthwith notify the American Arbitration Association
58 (hereinafter "association") of the need for a hearing and shall request

1 7 CHAP. 103

2

3 the association to provide to the commissioner forthwith a list of names
4 of persons chosen by the association from the association's panel of
5 labor arbitrators to potentially serve as hearing officers together with
6 relevant biographical information on each arbitrator. Upon receipt of
7 said list and biographical information, the commissioner [of education]
8 shall forthwith send a copy of both simultaneously to the employing
9 board and the employee. The commissioner shall also simultaneously
10 notify both the employing board and the employee of each potential hear-
11 ing officer's record in the last five cases of commencing and completing
12 hearings within the time periods prescribed in this section.

13 § 5. Paragraph c of subdivision 3 of section 3020-a of the education
14 law is amended by adding a new subparagraph (i-a) to read as follows:
15 (i-a)(A) Where charges of incompetence are brought based solely upon a
16 pattern of ineffective teaching or performance of a classroom teacher or
17 principal, as defined in section three thousand twelve-c of this arti-
18 cle, the hearing shall be conducted before and by a single hearing offi-
19 cer in an expedited hearing, which shall commence within seven days
20 after the pre-hearing conference and shall be completed within sixty
21 days after the pre-hearing conference. The hearing officer shall estab-
22 lish a hearing schedule at the pre-hearing conference to ensure that the
23 expedited hearing is completed within the required timeframes and to
24 ensure an equitable distribution of days between the employing board and
25 the charged employee. Notwithstanding any other law, rule or regulation
26 to the contrary, no adjournments may be granted that would extend the
27 hearing beyond such sixty days, except as authorized in this subpara-
28 graph. A hearing officer, upon request, may grant a limited and time
29 specific adjournment that would extend the hearing beyond such sixty
30 days if the hearing officer determines that the delay is attributable to
31 a circumstance or occurrence substantially beyond the control of the
32 requesting party and an injustice would result if the adjournment were
33 not granted.
34 (B) Such charges shall allege that the employing board has developed
35 and substantially implemented a teacher or principal improvement plan in
36 accordance with subdivision four of section three thousand twelve-c of
37 this article for the employee following the first evaluation in which
38 the employee was rated ineffective, and the immediately preceding evalu-
39 ation if the employee was rated developing. Notwithstanding any other
40 provision of law to the contrary, a pattern of ineffective teaching or
41 performance as defined in section three thousand twelve-c of this arti-
42 cle shall constitute very significant evidence of incompetence for
43 purposes of this section. Nothing in this subparagraph shall be
44 construed to limit the defenses which the employee may place before the
45 hearing officer in challenging the allegation of a pattern of ineffec-
46 tive teaching or performance.
47 (C) The commissioner shall annually inform all hearing officers who
48 have heard cases pursuant to this section during the preceding year that
49 the time periods prescribed in this subparagraph for conducting expe-
50 dited hearings are to be strictly followed. A record of continued fail-
51 ure to commence and complete expedited hearings within the time periods
52 prescribed in this subparagraph shall be considered grounds for the
53 commissioner to exclude such individual from the list of potential hear-
54 ing officers sent to the employing board and the employee for such expe-
55 dited hearings.
56 § 6. The education law is amended by adding a new section 211-e to
57 read as follows:

1 CHAP. 103 8

2

3 § 211-e. Educational partnership organizations. 1. The board of educa-
4 tion of a school district, and the chancellor of the city school
5 district of the city of New York, subject to the approval of the commis-
6 sioner, shall be authorized to contract, for a term of up to five years,
7 with an educational partnership organization pursuant to this section to
8 intervene in a school designated by the commissioner as a persistently
9 lowest-achieving school, consistent with federal requirements, or a
10 school under registration review.
11 2. Notwithstanding any other provision of law, rule or regulation to
12 the contrary, and except as otherwise provided in this section, such
13 contract shall contain provisions authorizing the educational partner-
14 ship organization to assume the powers and duties of the superintendent
15 of schools for purposes of implementing the educational program of the
16 school, including but not limited to, making recommendations to the
17 board of education on budgetary decisions, staffing population deci-
18 sions, student discipline decisions, decisions on curriculum and deter-
19 mining the daily schedule and school calendar, all of which recommenda-
20 tions shall be consistent with applicable collective bargaining
21 agreements. Such contract shall include district performance expecta-
22 tions and/or benchmarks for school operations and academic outcomes, and
23 failure to meet such expectations or benchmarks may be grounds for
24 termination of the contract prior to the expiration of its term. Such
25 contract shall also address the manner in which students will be
26 assigned to the school, the process for employees to transfer into the
27 school, the services that the district will provide to the school, and

28 the manner in which the school shall apply for and receive allocational
29 and competitive grants.
30 3. The board of education shall retain the ultimate decision-making
31 authority over the hiring, evaluating, termination, disciplining, grant-
32 ing of tenure, assignment of employees serving in the school as well as
33 with respect to staff development for those employees, together with
34 authority concerning all other terms and conditions of employment, all
35 of which decisions shall be made in a manner consistent with applicable
36 collective bargaining agreements. However, notwithstanding any law, rule
37 or regulation to the contrary, upon the effective date of the contract,
38 the educational partnership organization shall be authorized to exercise
39 all powers of a superintendent of schools with respect to such employ-
40 ment decisions, including but not limited to making recommendations, as
41 applicable, to the board of education in connection with and prior to
42 the board of education making decisions regarding staff assignments, the
43 hiring, the granting of tenure, the evaluating, the disciplining and
44 termination of employees, as well as concerning staff development. The
45 employees assigned to the school shall solely be in the employ of the
46 school district and shall retain their tenure rights and all other
47 employment rights conferred by law, and service in the school shall
48 constitute service to the school district for all purposes, including
49 but not limited to, the requirements for criminal history record checks
50 and participation in public retirement systems. Notwithstanding any
51 other provision of law to the contrary, for purposes of article fourteen
52 of the civil service law, employees in the school shall be public
53 employees of the school district as defined in subdivision seven of
54 section two hundred one of the civil service law and shall not be deemed
55 employees of the educational partnership organization by reason of the
56 powers granted to the educational partnership organization by this
57 section. All such employees shall be members of the applicable negotiat-
58 ing unit containing like titles or positions for the public school

1 9 CHAP. 103

2

3 district in which such school is located, and shall be covered by the
4 collective bargaining agreement covering that public school district's
5 negotiating unit, except that the duly recognized or certified collec-
6 tive bargaining representative for that negotiating unit may modify or
7 supplement, in writing, the collective bargaining agreement in consulta-
8 tion with the employees of the negotiating unit working in the school.
9 All such modifications of, or supplements to the collective bargaining
10 agreement are subject to ratification by the employees employed within
11 the school and by the board of education of the public school district,
12 consistent with article fourteen of the civil service law. Upon the
13 effective date of the school district's contract with the educational
14 partnership organization, the educational partnership organization shall
15 be empowered to make recommendations to the board of education with
16 respect to the scope of, and process for making modifications and addi-
17 tions to the collective bargaining agreement.

18 4. Where a recommendation is made by the educational partnership
19 organization to the board of education pursuant to subdivision two or
20 three of this section, and such recommendation is denied, the board of
21 education shall state its reasons for the denial, which shall include an
22 explanation of how such denial will promote improvement of student
23 achievement in the school and how such action is consistent with all
24 accountability plans approved by the commissioner for the school and the
25 school district. Nothing in this subdivision shall be construed to
26 prevent a board of education from denying a recommendation of the educa-
27 tional partnership organization based upon the board of education's
28 determination that carrying out such recommendation would result in a
29 violation of law or violation of the terms of an applicable collective
30 bargaining agreement. If the board of education rejects a recommendation
31 of the educational partnership organization to terminate a probationary
32 employee assigned to the school or to deny tenure to an employee
33 assigned to the school, it shall be the duty of the board of education
34 to transfer such employee to another position in the school district
35 within such employee's tenure area for which the employee is qualified,
36 or to create such a position.

37 5. For purposes of this section the following terms shall have the
38 following meanings:

39 (i) "educational partnership organization" means a board of cooper-
40 ative educational services, a public or independent, non-profit institu-
41 tion of higher education, a cultural institution, or a private, non-pro-

42 fit organization with a proven record of success in intervening in
43 low-performing schools, as determined by the commissioner, provided that
44 such term shall not include a charter school;
45 (ii) "board of education" means the trustees or board of education of
46 a school district, or, in the case of a city school district of a city
47 having a population of one million or more, the chancellor of such city
48 district;
49 (iii) "school district" means a common, union free, central, central
50 high school or city school district, other than a special act school
51 district as defined in section four thousand one of this chapter.
52 (iv) "superintendent of schools" means the superintendent of schools
53 of a school district, and, in the case of a city school district of a
54 city having a population of one million or more, a community superinten-
55 dent and the chancellor of such city district when acting in the role of
56 a superintendent of schools.
57 § 7. This act shall take effect immediately; provided however that the
58 provisions of sections one, two, three, four and five of this act shall

1 CHAP. 103 10

2

3 take effect July 1, 2010, provided, further, if this act shall become a
4 law after such date it shall take effect immediately and shall be deemed
5 to have been in full force and effect on and after July 1, 2010.

6

7 The Legislature of the STATE OF NEW YORK ss:

8 Pursuant to the authority vested in us by section 70-b of the Public
9 Officers Law, we hereby jointly certify that this slip copy of this
10 session law was printed under our direction and, in accordance with such
11 section, is entitled to be read into evidence.

12

13 MALCOLM A. SMITH SHELDON SILVER

14 Temporary President of the Senate Speaker of the Assembly

Appendix E—Synthesis Document of the Locally-Selected Measures Workgroup from January 2011 Meeting

****Note - this report reflects the thinking of the group early in the process, prior to reporting-out to the full Task Force****

The following categories of options were discussed. (More detailed description in appendix)

- A. Currently-available commercially-available assessments
- B. District/BOCES issue RFP for assessment development
- C. District/BOCES developed assessments
- D. Group or team metric
- E. Teacher-created assessments

Discussion of Options

- Through discussion of the above options, the workgroup came to the consensus that options A, B, and C are the preferred options.
- The workgroup feels that D is an option, but not likely to be selected for ELA and Math 3 – 8 since most districts are likely to have alternatives. This will be discussed in more depth when non-tested subjects are addressed.
- The workgroup strongly feels that under no circumstances, should the State exams be used for the local option. They feel that this would place too much emphasis on the state tests and have a **profoundly negative effect on instruction**. Further, many group members feel that the use of state tests for the local options is in contrast with the language of the law.

The group could not come to a consensus on an option for teacher/principal agreements (i.e., school-created assessments). Many in the group believe that a process can be designed to be rigorous and comparable given appropriate guidelines/rubrics from the state on definitions for rigorous and comparable. Others in the group disagree with the use of this option, questioning its ability to be comparable across a school district. The group has received feedback from the Commissioner that school-selected assessments should not be an option. At this point in time, the workgroup feels that schools should not be precluding from using teacher-created assessments as an option.

Institute **currently-available commercial assessments**, i.e., off-the-shelf assessments from vendors from menu of approved options based on SED RFQ.

Instruments Could Include (subject to RFQ):

- *NWEA MAP*
- *Scantron Performance Series*
- *CTB Acuity*
- *ACT Explore / College Board Readiness*

Issue RFP for vendor-built custom assessments. Would require an RFP process

and ongoing vendor relationship management (both fiscal and technical). The RFP would need to conform to SED parameters. SED’s parameters would need to address comparability, validity, rigor, degree to which test can be aligned to state standards, and degree to which test must drive desirable classroom practice.

Allow **Districts and/or BOCES** to design and develop their own local assessment system. These assessments would be standardized across the district and/or BOCES and would need to meet requirements / parameters defined by SED. SED’s parameters would need to address comparability, validity, rigor, degree to which test can be aligned to state standards, and degree to which test must match classroom instruction.

Use a **group or team metric**. A district may decide to use a school-wide or grade-wide growth metric as the local assessment option to, for example, reduce individual competitiveness and emphasize team-based teaching. For example, a district may decide to use the percent of a particular subgroup who moved from 1 to 2 (calculated school-wide) as a portion of the local measure.

Work-group endorsed options:

Based on previous workgroup discussion and staff research and analysis, the following is presented as highly preliminary assessment of the options the group feels should be considered as local assessment options with A, B , C, and E as most appropriate for grades 4-8 ELA and Math.

- Strong:** Strong comparability, strong alignment (after augmentation), meets industry standards for reliability and validity for high-stakes assessments, positive impact on instruction, cost-effective and feasible.
- Partial**
- Poor / Minimal**

<i>Criteria</i>	A. Institute Currently-Available Non-SED Assessments (e.g., off-the-shelf tests from vendors)	B. Institute Vendor-Built Custom Tests	C. Districts / BOCES Build Their Own Local Assessments	D. Use as Group or Team Metric	E. Teacher-Created Assessments
Comparability	<input checked="" type="radio"/> Assume decision for this option is district/BOCES-wide.	<input checked="" type="radio"/> Assume decision for this option is district/BOCES-wide.	<input checked="" type="radio"/> Assume decision for this option is district/BOCES-wide.	<input checked="" type="radio"/> Fully comparable since uses state metrics and would compare groups to same groups (schools, grades, etc) across the district or state.	<input checked="" type="radio"/> Could be comparable if SED provides rubrics to ensure comparability
Alignment to State Standards	<input checked="" type="radio"/> Would require extensive test augmentation requiring resources familiar with NYS Standards	<input checked="" type="radio"/> Will be designed from outset to be aligned.	<input checked="" type="radio"/> Will be designed from outset to be aligned.	<input checked="" type="radio"/>	<input checked="" type="radio"/> Will be designed from the outset to be aligned

Reliability and Validity	<input checked="" type="radio"/> Likely issue with tails (unless Computer adaptive testing is used) Most available tests too general - Cannot make specific inferences about student ability in terms of what the student was taught.	<input type="radio"/> Will be designed from outset to meet industry standards.	<input type="radio"/> Will be designed from outset to meet industry standards.	<input type="radio"/>	<input checked="" type="radio"/> Could be valid and reliable if SED provides rubrics to ensure reliability and validity
Effect on Instruction	<input type="radio"/> Could be better if vendor has tools that connect scores to curricular resources	<input type="radio"/> Customization to ensure desired impact on classroom practice.	<input type="radio"/> Customization to ensure desired impact on classroom practice.	<input type="radio"/> Possibly mitigates in-school competition. Or reduces individual accountability.	<input type="radio"/> Assessments are specifically aligned to classroom instruction and therefore will have a positive effect.
Cost-Efficiency and Feasibility	<input checked="" type="radio"/> Potentially costly; technology issues could prevent implementation	<input type="radio"/> Could be expensive. Development timeline could be too lengthy.	<input type="radio"/> Could be expensive. Development timeline could be too lengthy.	<input type="radio"/> No new tests.	<input checked="" type="radio"/> Could be costly in terms of teacher and principal training, time for approval/agreement process. The feasibility of these assessments will be different from district to district.
Timing	<input checked="" type="radio"/> Could be challenging to work with vendor to ensure scores available on time.	<input type="radio"/> Because we develop them, we set timing.	<input type="radio"/> Because Districts / BOCES develop them, they set timing.	<input type="radio"/> Same timing as state tests.	<input type="radio"/>

Appendix F—Workgroup Opinion on State Tests for 20% Local

The law states: “Forty percent of the composite score of effectiveness shall be based on student achievement measures as follows:

(i) twenty percent of the evaluation shall be based upon student growth data on state assessments as prescribed by the commissioner or a comparable measure of student growth if such growth data is not available; and (ii) twenty percent shall be based on other locally selected measures of student achievement that are determined to be rigorous and comparable across classrooms in accordance with the regulations of the commissioner and as are developed locally in a manner consistent with procedures negotiated pursuant to the requirements of article fourteen of the civil service law.”

The intent of the law was to have 20% growth based on state assessments and 20% locally selected multiple measures of student achievement based on other measures beyond state assessment or the measure of student growth being used for non-tested. It seems that the language is clear that something other than state assessments should be used in this locally selected 20%. The workgroup indicated that they felt that not all districts would have alternative assessments to use for this portion of the evaluation. If you take into account the list of possible “other” measures below it is a far reach to say these measures don’t currently exist and if they don’t already meet the state determined rigor criteria and align with standards they could be modified to do so.

There is also concern that the 20% locally developed workgroup has recommended the possibility of a single measure as a viable option in its assumptions.

Multiple-measure systems improve the accuracy and stability of teachers’ evaluations by reducing reliance on any single measure of a teacher’s performance, while also reducing the likelihood that teachers will engage in excessive test preparation or other forms of test-focused instruction (Booher-Jennings, 2005; Hamilton et al., 2007; Stecher et al., 2008).

The use of multiple measures in teacher evaluation systems creates a system that provides timely and relevant information that can guide and enhance professional development throughout the year and throughout a teacher’s career – (Resource AFT A Guide for Developing Multiple Measures for Teacher Development and Evaluation). Evidence of student learning should not be limited to standardized tests. This local option should be aligned with NY State Standards/Common Core meet statewide criteria and consist of multiple measures of student performance such as:

- Criterion referenced test
- Curriculum based assessments
- Formative assessments
- Norm referenced tests
- Performance assessments
- Portfolio or student work

- Summative assessments

I am concerned that there might be a press from some fronts to reconsider the idea of using the state tests as the “other 20%” as I have heard that possibility uttered in other presentations from state folks. For the ease and expediency of implementation, I understand it would be a tempting option. I am glad our committee has resisted so strongly. And, that we are standing behind the perceived intent of the law and also the ideals of authentic assessments as indicators of student learning.

Appendix G—Scenarios for Use of State Test Results in the Other 80%, by member of Metrics Workgroup, March 24, 2011

The below table is intended to help illustrate the different scenarios where state tests *may or may not* be used for the “other 80%” of a principal or a teacher’s evaluation. This is not intended as an argument for or against these options but rather to help illustrate what the scenarios are.

In addition, a key question that runs through all these scenarios is whether this should be provided as a choice for principal and teachers – not a requirement - or whether they should be explicitly excluded as an option. It would seem from the local assessment group’s memo that they believe all of these options should be explicitly not be permissible, while other Task Force members felt that some or all of these scenarios could be permissible.

Scenarios	Comments
<p>Exact same measure is used for 20% state growth & 20% local</p> <p>i.e., 7th grade teacher’s Math “20 state growth” score is used for 7th grade teacher’s math “20 state growth” score</p>	<p>This would appear to be the biggest source of controversy –with some members of Task Force arguing strongly that the same measure should not be counted twice (a.k.a. as “double dip” or “double jeopardy”)</p>
<p>Same test but different measure used for 20% state growth</p> <p>e.g., 7th grade teacher has a math “20% state growth” score based upon state’s growth model; uses any of the following for local measure – performance (ex. % proficient) on 7th grade math state tests, performance or growth on 7th grade math tests for specific population of students (ex. Special education students), progress toward goal on 7th grade math tests (ex. increase of x% of students who move from level 2 to level 3)</p>	<p>This appeared to gain a bit more support among some members of Task Force. For example, it was felt by some Task Force members that for principals in particular this would be a viable option since they may, for example, have a growth score only based on math and ELA g4-8 growth but want their local score to be based on more school-specific goals.</p>
<p>Different tests than used for 20% state growth (or teacher/ principal does not currently have a 20% state calculated growth score)</p> <p>e.g., 4th grade teacher has a math and ELA “20% state test growth” score & uses 4th grade science score as local measure; teacher of a course leading to a Regents</p>	<p>A number of Task Force members felt that this was acceptable – in particular, it felt odd to not allow for example a US History teacher to have the US History Regents exam count towards any part of their evaluation</p>

<p>exam or 4th/ 8th grade science – currently has no state-calculated growth score for “20% state growth”; principal’s evaluation allows for use of Regents exam in their local or 60% other measures when not used for 20% state growth score</p>	
<p>Principal evaluation only – allowable measure for 60% other measures could include measure of retention of effective teachers or some other measure of teacher quality in their building (as an option, not necessarily a requirement, given that it was felt that some principals do not have as much “control” over teacher quality as others)</p>	<p>Again, it was felt by some Task Force members that depending on how this was measured this could be a permissible measure. For example, including a measure like “maintained or increased % of H or E teachers in building” was not duplicative with that principals own state test growth score.</p>

Appendix H—Teacher-Created Assessments, by member of Workgroup on Locally-Selected Measures, Jan 30, 2011

Introduction: In group discussions, the possibility of allowing a Teacher-Created Assessment option on the menu for Locally Created or Developed Assessments has been raised. The list below attempts to articulate some of the pros, cons, and neutral considerations for this option.

Pros:

Allows teachers to:

- customize assessments to match the unique make-up of their student populations, such as ESL or SWD cohorts.
- craft creative and higher-order thinking assessments beyond the standardized test format- such as performance-based assessments.
- address district and building-specific learning goals.
- effectively measure the degree to which students have mastered what is actually being taught.
- use relevant and meaningful data to direct classroom curriculum and instructional methods.
- supplement material covered on standardized tests to decrease the likelihood of test-centered curriculums.
- gain professional development. Working with evaluators to ensure that assessments meet NYS standards will provide teachers with valuable on-the-job training. As all educators are required to demonstrate competency in designing high-quality assessment tools as part of their teacher preparation program and certification process, this training would serve to further supplement a skill set already in place. Such professional development would be particularly meaningful to newly hired staff members.
- feel more empowered with regard to the student achievement piece of their evaluation. Evaluator feedback would inform the instructor and, in turn, afford the evaluator greater insight into what is actually being taught and assessed in the classroom.
- update the curriculum and local assessment tool as needed. (e.g. -An ELA instructor might wish to use new literature, possibly necessitating a change in the local assessment tool. If the assessment were teacher-created, such a change would be easy to make.

- save their schools money as the district would not have to pay vendors for tests or hire substitute teachers for release time to allow groups of teachers to design a "one-size-fits-all document".

Neutral Comments:

- Many schools already require that teachers submit their final course examination to the administrator. If the 20% Teacher-Created Local Assessment was based on student achievement, then such administrators would be evaluating assessments they've already seen or are used to seeing in the course of their regular duties. In other words, this would not be a significant amount of extra work for those administrators that already preview the final exams given in their schools.

- As the Teacher-Created Option would be one of several options on the menu, schools would not be under any obligation to choose it. However, by allowing it on the menu, districts would have the opportunity to take advantage of the above listed benefits should it be locally negotiated and determined to meet the needs of the district.

- The New York State Education Department would craft a rubric that would measure the degree to which an assessment is rigorous, valid, reliable and meets the core standards. This rubric would be used for all assessments, allowing for comparability in each of the categories. The law does not currently specify that comparability means the same assessment, leaving open the possibility that different assessments could be comparable if they achieve the same score using the SED rubric.

Cons:

- Additional time would be required of administrators to evaluate Teacher-Created Assessments. Administrators have ever increasing demands on the time and responsibilities, including their new Principal Evaluation requirements. One way to aid over-worked administrators would be to permit assistant principals or teacher leaders, such as certified consultant teachers or department representatives to evaluate the assessment tool using the SED rubric. The administrator might even have a list of struggling teachers or new hires that he or she would need to look at personally in terms of their assessments- but permit other, highly effective teachers to submit their assessments to teacher leaders. As not all districts would be comfortable with such arrangements, this possibility might be a locally negotiated item.

Another way to lessen the time burden would be to have the administrator evaluate the assessment tool during the pre-evaluation meeting with the teacher. As most districts required such a meeting prior to the first classroom observation, the extra time to look over the Teacher-Created Assessment would not be quite as onerous as scheduling a separate meeting.

It is noteworthy, however, that all educational professionals- be they administrators or instructors- are being asked to take on additional responsibilities in the interest of

providing quality teacher and principal evaluations- an important step in improving student achievement.

- Subjectivity may be a factor when using the SED-created rubric to evaluate Teacher-Created Assessments. However, if more than one person were to independently evaluate an assessment and compare results, such as an assistant principal, certified consultant teacher, or department representative, along with the evaluator, the chances of unintended subjectivity would lessen and give the evaluators more confidence that their determination is corroborated by other trained and experienced professionals.

- Teachers might need release time to create a local assessment. However, an individual teacher might require less time, if at all, than a subject or grade-specific group of instructors crafting a district-wide document. If release time was required, it would most likely cost less than purchasing outside vendor exams for the entire grade or subject area. Creating assessments throughout the year is part of an instructor's job expectations. Many teachers may already have assessments in place that would meet state standards. If needed, release time could be a locally negotiated item.

***Appendix I—Local Bargaining Issues, by one member of Workgroup
on Locally-Selected Measures***

Starts on next page.



Commission on
Independent
Colleges and
Universities

17 Elk Street 518-436-4781 www.cicu.org
PO Box 7289 518-436-0417 fax www.nycolleges.org
Albany, NY 12224 mail@cicu.org

Quality
Access
Diversity
Choice

To: Members of the Regents Task Force on Teacher and Principal
Effectiveness Committee on Locally Selected Assessments

From: Elizabeth Van Nest, Vice President for Legal Affairs and General Counsel

Re: Analysis of Underlying Law (Chapter 103, Laws of 2010) as it Relates to
Locally Selected/Developed Assessments and Collective Bargaining

Date: January 28, 2011

Background

The following analysis is provided in response to the State Education Department staff's request during our Committee's January 26 conference call that I summarize in writing various points that I had raised during the conference call, in which not all of us were able to participate. The points that I raised were in response to a suggestion by a member of our Committee that the Committee consider concluding that only the procedures relating to the locally developed assessments, and not their content, should be the subject of collective bargaining, to which I had responded that the language of the new law must govern any conclusions, recommendations, or parameters that we might provide to the Department that could be used as the basis for regulations based on the new law.

I further stated that if there were ambiguity in that language, the Committee should first have clarity from an authoritative source on the intent of the language of the new law before proceeding, since regulations inconsistent with the law would be subject to challenge, and especially since our Committee has been discussing and providing options for the local selection and development of assessments that include currently available non-State Education Department assessments, custom vendor-built tests, local assessments built by school districts and BOCES, and the use of group or team metrics.

Analysis of Applicable Collective Bargaining and the New Law

In taking a closer look at the entirety of the new law underlying our Task Force's and its various Committees' activity, I note that the collective bargaining process appears several times, i.e., first in relation to "locally developed procedures" subject to collective bargaining, with regard to the annual professional performance reviews as they may relate to, among other things, promotion, retention, tenure determination, termination and supplemental compensation, as well as teacher and principal development. Additionally, the new law states that: "**Except for the student growth measures" (citing specific portions of the new law, specifically paragraphs e, f, and g of Subdivision 2 of Section 3012-C of the Education Law), "the elements comprising the composite effectiveness score shall be locally developed, consistent with the standards prescribed in the regulations of the Commissioner, through negotiations conducted pursuant to the requirements of Article 14 of the Civil Service Law" (collectively bargained).**

The New Law as it Relates to the Charge to the Committee

The excluded paragraphs above are paragraphs that specifically govern the subject-matter with which our Committee is charged, and each of those paragraphs contain essentially the same language except where each refers to a specific time period and percentage of the total 40% student growth-based portion of the annual professional performance review, of which the locally selected/locally developed measures are a part. Therefore, I have set forth only Paragraph e below, with the language at issue in our January 26 conference call in bold text, that language being identical in each of the excluded paragraphs (e, f, and g) referred to above:

“e. For annual professional performance reviews conducted in accordance with paragraph b of this subdivision in the two thousand eleven--two thousand twelve school year, forty percent of the composite score of effectiveness shall be based on student achievement measures as follows:

- (i) Twenty percent of the evaluation shall be based upon student growth data on state assessments as prescribed by the commissioner or a comparable measure of student growth if such growth data is not available; and
- (ii) **Twenty percent shall be based on other locally selected measures of student achievement that are determined to be rigorous and comparable across classrooms in accordance with the regulations of the commissioner and as are developed locally in a manner consistent with procedures negotiated pursuant to the requirements of article fourteen of the civil service law.” (emphasis added)**

The first issue presented is what are the parameters that the assessments must meet, according to the language in the new law? It appears from the language that there are two:

1. Twenty percent shall be based on other locally selected measures of student achievement that are determined to be rigorous and comparable across classrooms in accordance with the regulations of the commissioner;

AND

2. [A]s are developed locally in a manner consistent with procedures negotiated pursuant to the requirements of article fourteen of the civil service law.” (emphases added)

As indicated above, the suggestion has been made that the Committee should conclude that only the procedures for developing the local student growth/achievement assessments should be the subject of collective bargaining, and not the determination of the selection of the local assessments. Our Committee has discussed and included in its consensus of recommended options the options of local assessments that are either district or BOCES built, or those that are developed through a group or team metric, consistent with the language that assessments to fulfill this portion of the law are to be “locally developed.”

The ambiguity emerges because the language providing for local development of assessments implies that more than just the procedures for their development is to be the subject of collective bargaining, which then leads to the question of just what the language that includes “procedures negotiated pursuant to [collective bargaining]” means. For instance, does it mean, as has been suggested for the Committee’s consideration that only the procedures themselves are to be negotiated, or does it mean that the “procedures” that are to be collectively bargained would or could include not only how the local assessments are to be developed, but their actual development? These alternatives are important ones, because they cut to the heart of how inclusive the developmental process is to be at the local level, and who is to be included.

Admittedly, these are questions of substantive import, and yet the form of the language must govern the conclusions that are drawn from it. Since there is apparent ambiguity in the language, there is also need for clarity to ensure that the Committee’s and the Task Force’s recommendations— and ultimately regulations— are consistent with both the letter and the intent of the new law.

Collective Bargaining and the Remaining 60% of the Evaluations

The new law is clearer with regard to the remaining 60% non-growth measures on which the annual professional performance reviews are to be based (Education Law, § 3012-c(2)(h)), as follows:

“h. The remaining percent of the evaluations, ratings and effectiveness scores shall be locally developed, consistent with the standards prescribed in the regulations of the commissioner, through negotiations conducted pursuant to article fourteen of the civil service law.”

Conclusion

As a practical matter, absent authoritative guidance, such as perhaps: 1) a clarification of the operative legislative intent; 2) an opinion from GOER, the Governor’s Office of Employee Relations, the office that deals with collective bargaining issues; or 3) ultimately a chapter amendment, if necessary, to clarify the meaning of the language, the Committee’s work is severely hampered because the parameters under which it is to proceed on key substantive matters are not clear.

This lack of clarity pertains to key issues with which the Committee has been charged under the law to advise the Commissioner, and as a member of the Committee, an attorney, and admittedly, one that is not an expert in collective bargaining, I am not comfortable in proceeding with these important issues absent further clarity.

Appendix J—Proposed Framework for Principal Evaluation

Measures for 60% Principal Evaluation	Weight	% of total	Score
Broad assessment of principal leadership and management actions which <ul style="list-style-type: none"> • are based on multiple measures; • include feedback from the supervisor plus two additional stakeholders; and • include at least two sources of evidence for each measure, with feedback being one source of evidence. 	<i>Locally determined</i>	<i>Locally determined</i>	<i>Locally determined</i>
Assessment of principal's skill in supervising and developing teachers			
Progress against ambitious, measurable school academic or learning environment goals			
Progress against individual professional growth goals			
TOTAL	<ul style="list-style-type: none"> • The numeric score will reflect the total points earned out of the total possible points allowed (which is dependent on the evaluation tool utilized). • This number should not be expressed as 1-4 to correlate with HEDI ratings, for the purposes of creating a composite score (although this is yet to be fully determined). • The full range of available points should be possible; this rating should not be an “all or nothing” amount. (For example, if a 100-point scale were used for the composite scale, 60 points would be available in this measure.). • Districts will have to make public how points will be assigned based on 60% measures. 		

Appendix K—60% Principals Summary Table of Proposed Elements and Measures

Element	Required?	Potential Requirements/Guidance (Subject to Local Bargaining)
<p>Broad Assessment of Leadership and Management Actions</p> <ul style="list-style-type: none"> • to sustain and/or improve instruction and teaching* • to impact the conditions for learning • to engage parents and community in support of student learning <p><i>* Includes school-wide actions and actions to improve curriculum and assessment, whereas the 3rd element (below) is more specifically focused on the Principal's skill in supervising and developing teachers</i></p>	<p><i>Required as a major part of the 60%</i></p>	<p>Required: Supervisor review of principal behaviors, skills, and accomplishments using a standardized rubric (the Workgroup hasn't yet developed a list of criteria that such rubrics would have to meet, but it is envisioned that such rubrics could be 3rd-party-developed or district-developed)</p> <p>Required: Formalized and standardized feedback (through surveys, meetings, etc.) from at least two of the following additional sources: teachers, students, parents/families, community stakeholders, self-assessment.</p> <p>The supervisor could also take into consideration some or all of the following optional sources of evidence:</p> <ul style="list-style-type: none"> • Progress against ambitious, measurable, school-wide goals set jointly with supervisor to address substantive issues identified through analysis of data (<i>Notes: This is not intended to duplicate the 40% student achievement metrics. Some members of the Workgroup felt that this should be a separate element. At our last meeting that issue was not clearly decided.</i>) • School site visits or observations (e.g., Tri-State assessment visits, NYC Quality Review, or other types of visits) to observe systems and processes on use of data, monitoring curriculum and instruction, and programs and services • Review of school documents, systems, and records (e.g., budgets, safety data, compliance records) to see that they are in compliance, well understood by staff, and used to meet school learning goals. • Other sources as determined locally <p>Workgroup members offered examples of other sources of evidence and the information that could be gathered:</p> <ul style="list-style-type: none"> • Observations of meetings, professional development planning/programs, other school improvement work • Observations of school facilities and technology are maximized to support student learning • No material violations of law, regulations, or district guidelines occur; those that occur are remedied. • School visits show that regular systems for recording and analyzing student learning data inform school decision making, particularly the assignment of staff, allocation of resources, curriculum implementation and development, lesson planning and professional development • Family and community members provide tangible and intangible support of school goals • Family participation rates for specific events • Surveys show staff, family and student understanding and support of student learning goals • Community engagement plans and strategies • Participation of key stakeholders in support of the school and student learning • Maximize time spent on quality instruction • Create a personalized and motivating learning environment for students • Create a comprehensive, rigorous and coherent curricular program • Promote and protect the welfare and safety of students and staff • Obtain, allocate, align and efficiently use human, fiscal and technological resources

Summary Table (continued)

Element	Required?	Potential Requirements/Guidance (Subject to Local Bargaining)
<p>Principal’s Skill in Managing School Operations and Resources <i>(This element was not fully discussed by Workgroup)</i></p>	<p><i>Some Workgroup members favored including this as a separate element.</i></p>	<p>In assessing this element, the supervisor could take into consideration some or all of the following optional sources of evidence:</p> <ul style="list-style-type: none"> ● School site visits or observations ● Formalized and standardized feedback (through surveys, meetings, etc.) from sources such as parent/family, teacher, student, community stakeholders ● Progress against ambitious, measurable, school-wide goals set jointly with supervisor to address substantive issues identified through analysis of data ● Review of school documents, systems, and records (e.g., budgets, safety data, compliance records) ● Other sources as determined locally
<p>Assessment of Skill in Supervising and Developing Teachers <i>This is a major responsibility of principals, and all principals have some influence on teacher effectiveness. Therefore, principals should be accountable for exercising the full extent of their authority over teachers—whether that includes hiring, professional development, granting or denying tenure, compensation, or firing.</i></p>	<p><i>Required</i></p>	<p>Depending on local context, a principal’s influence may be limited with regard to the principal’s ability to hire, fire, grant or deny tenure, compensate and retain teachers. Therefore, measures should be carefully selected to ensure fairness and efficacy, and districts should have the flexibility to select appropriate metrics, depending on principal authority. Examples:</p> <ul style="list-style-type: none"> ● Quality/effectiveness of teacher evaluations: <ul style="list-style-type: none"> ○ Processes focus on teacher strengths, challenges, & effectiveness in meeting student learning goals ○ Timely observations and feedback ○ Effective differentiation of teachers, consistent with student outcome (achievement/growth) data ○ Use of Teacher Improvement Plans ○ Retention, tenure, and salary decisions consistent with evaluation/effectiveness ● Rates of recruitment and retention of teachers rated effective or highly effective ● Rates of teacher attendance ● Records showing that ineffective staff are dismissed after given a fair opportunity to improve ● Skill in improving teacher effectiveness over time <ul style="list-style-type: none"> ○ Improvements in individual teachers’ student growth outcomes over the period that the teacher has spent under the principal’s leadership (<i>One Workgroup member voiced support for including this metric, while others expressed concern that it would lead to double-counting of student achievement or that it would be psychometrically noisy</i>) ● Provision of opportunities for teachers to collaborate with colleagues ● Provision of opportunities for accomplished teachers to share their expertise and serve as coaches, mentors, etc.

Summary Table (continued)

Element	Required?	Potential Requirements/Guidance (Subject to Local Bargaining)
<p>Progress Against School-Wide Goals</p>	<p><i>Some members of the Workgroup felt that this should be a separate element. At our last meeting that issue was not clearly decided.</i></p>	<ul style="list-style-type: none"> ● Goals must be selected jointly with the supervisor ● Goals must be ambitious and measurable ● Goals must address substantive issues identified through analysis of data (<i>Note: this is not intended to duplicate the 40% student achievement metrics</i>)
<p>Progress Against Individual Professional Development Goals</p>	<p><i>Many Workgroup members feel this should be Required.</i></p> <p><i>One Workgroup member feels this should be Optional.</i></p>	<ul style="list-style-type: none"> ● Goals must be grounded in clearly articulated domains and standards such as the ISLLC Standards or other leadership standards or professional development standards that are research-based and that directly link to and promote student achievement outcomes. ● Goals must be selected by the educational leader and shared with his/her supervisor. One Workgroup members stated that the selection of goals should be subject to the supervisor's approval, and that the supervisor should be able to supplement the selected goals to focus on areas of concern, weakness, or district priority. ● The educational leader and his/her supervisor will determine how acquisition of knowledge and skills will be demonstrated. ● The educational leader must provide evidence demonstrating acquisition, enhancement, and application of skills/expertise to his/her supervisor. <p><i>Many Workgroup members voiced strong support for the inclusion of professional development goals as an evaluation element and argued that if this were not included, districts would not take principal professional growth as seriously for all principals.</i></p> <p><i>Although there is broad agreement that an important part of any evaluation system is to inform and support professional development, one Workgroup member voiced disagreement regarding whether the professional development goals should be part of the performance management PROCESS vs. an evaluative measurement TOOL. That member expressed concern (a) that the inclusion of professional development goals will not support a system that effectively differentiates performance and (b) that basing the accountability rating on formative goal-setting could create incentives for the leader to select easier goals, thereby undermining their usefulness for both evaluative and formative purposes.</i></p>