MERA Conference

Teacher/Administrator Evaluation and Student Growth: Wayne-Westland Community Schools Model

Paul Salah, Ed.D.

Agenda

- A Brief History Lesson
- We had to face reality
- Wayne Westland's Model
- Sample Rubrics
- Building Administrator Evaluation

Learning Target

- I can verbally describe the Wayne-Westland Teacher evaluation model to a colleague using a few examples from the presentation.

Brief History Lesson

- January 4, 2010—District's required to adopt, minimally, an annual performance evaluation system for all teachers and administrators that:
  - Was **rigorous, transparent, and fair**;
  - Provided **timely and constructive feedback**; and
  - **Used multiple rating categories** that take into account data on **student growth as a significant factor**.
  - Was to be measured by **national, state or local assessments and other objective criteria**.
    - NOTE: Percentages of student growth, as a measure of teacher evaluation, **were not defined in January 2010**.

A Brief History Lesson

- **Districts had been working to solve the evaluation quandary, defining significant and the legislature ended up doing the work for districts.**
- June 2011—Section 1249 was further amended to add clarification to the January 2010 legislation.
  - **25% by 2013-2014**
  - **40% by 2014-2015**
  - **50% by 2015-2016**

Evaluation Drives Improvement

- Evaluations are not punitive
  - Professional Growth Plan
  - Not a "Gotcha!" trap!
- Evaluations should drive instruction and professional development decisions.

Taking Care of Two Birds with One Stone

- Faced with compliance, we asked ourselves, **how can we accomplish our primary objectives, as a district, without adding one more unrelated initiative**. How can we:
  - Support and refine good quality teaching and learning
  - Support the implementation of Professional Learning Communities
  - Improve student achievement
  - Support the School Improvement Process

We had to face reality

- "In Denver schools that did not make AYP, more than 98% of tenured teachers received the highest rating—satisfactory."—Widget Effect (2009)
- "On average, over the last three years, only 10% of failing schools issued at least one unsatisfactory rating to a tenured teacher." –Widget Effect(2009)

We had to face reality

- "Research on attitudinal change has long found that most of us change our behaviors somewhat before we get insights into new beliefs. The implication for approaching new change is clear. **Do not load up on vision, evidence and sense of urgency. Rather, give people new experiences in relatively nonthreatening circumstances and build on it, especially through interaction with trusted peers—very simple but hard to do when you are impatient for buy in."**—Fullan, M. (2010) *Motion Leadership, The Skinny to Becoming Change Savvy*.

We had to face reality

- "The debate about "value added" measures of teaching may be the most **divisive topic in teacher-quality policy today**. It has **generated sharp-tonged exchanges** in public forums, in news stories, and on editor pages. And it has **produced enough policy briefs to fell whole forests**."—Sawchuk, S. (2011) *Wanted: Ways to Assess the Majority of Teachers.* Education Week.

We had to face reality

- "Learning, broken down to its simplest equation, is ***change*; the brain must make a chemical transformation in the process of learning and a specific series of structural modifications to produce a new behavior**."—Sanchez, H. (2008) *A Brain Based Approach to Closing the Achievement Gap.*
  - Thus, the introduction of new concepts or changes are almost always initially rejected.

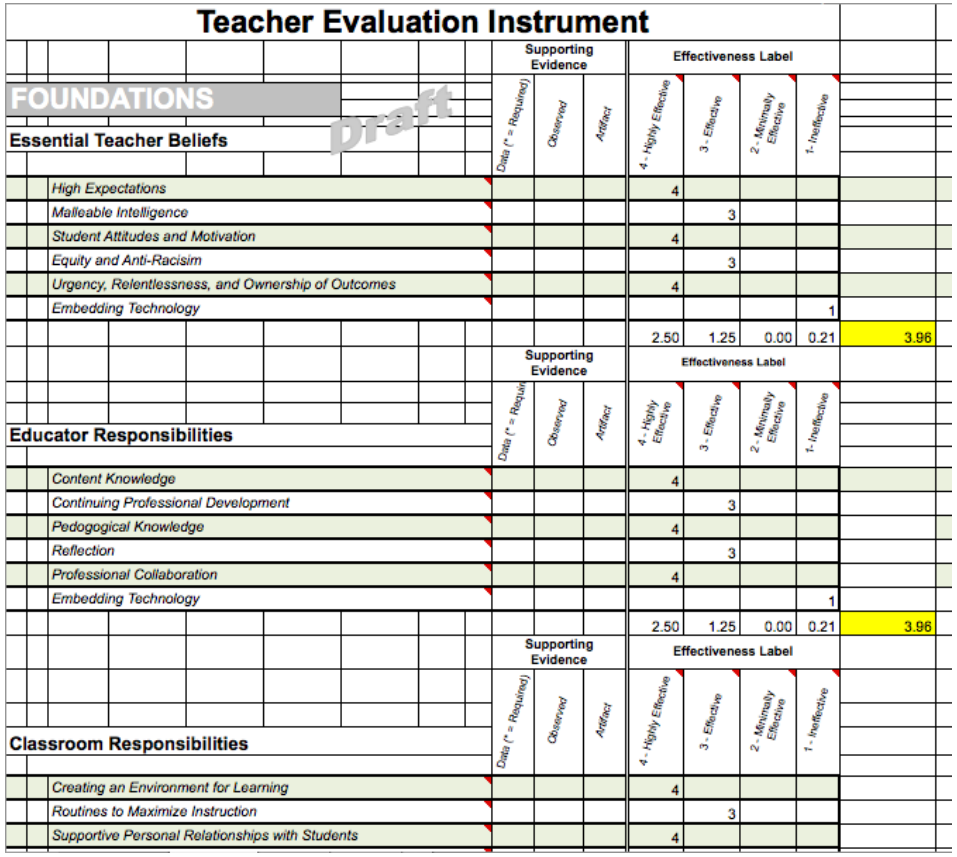
Wayne-Westland Evaluation Model

- Wayne-Westland contemplated two primary considerations related to the development of teacher evaluation:
  - Qualitative Measures
  - Quantitative Measures

Qualitative Measures

- Qualitative Measures:
  - 14 sections, each worth 5 points for a total of 70 points of the total evaluation.
  - Qualitative sections are **observable** and often **behavior oriented** items or characteristics.
  - In many instances the qualitative sections deal with **affective qualities** that are **observable** or **provable through empirical observation**, e.g. essential teacher beliefs.

Checklist



Teacher Evaluation Instrument

Draft

FOUNDATIONS

Essential Teacher Beliefs

	Supporting Evidence			Effectiveness Label				
	Data (* = Required)	Observed	Artifact	4 - Highly Effective	3 - Effective	2 - Minimally Effective	1- Ineffective	
High Expectations				4				
Malleable Intelligence					3			
Student Attitudes and Motivation				4				
Equity and Anti-Racisim					3			
Urgency, Relentlessness, and Ownership of Outcomes				4				
Embedding Technology							1	
				2.50	1.25	0.00	0.21	3.96

Educator Responsibilities

	Supporting Evidence			Effectiveness Label				
	Data (* = Requir	Observed	Artifact	4 - Highly Effective	3 - Effective	2 - Minimally Effective	1- Ineffective	
Content Knowledge				4				
Continuing Professional Development					3			
Pedogogical Knowledge				4				
Reflection					3			
Professional Collaboration				4				
Embedding Technology							1	
				2.50	1.25	0.00	0.21	3.96

Classroom Responsibilities

	Supporting Evidence			Effectiveness Label				
	Data (* = Required)	Observed	Artifact	4 - Highly Effective	3 - Effective	2 - Minimally Effective	1 - Ineffective	
Creating an Environment for Learning				4				
Routines to Maximize Instruction					3			
Supportive Personal Relationships with Students				4				

Quantitative Measures

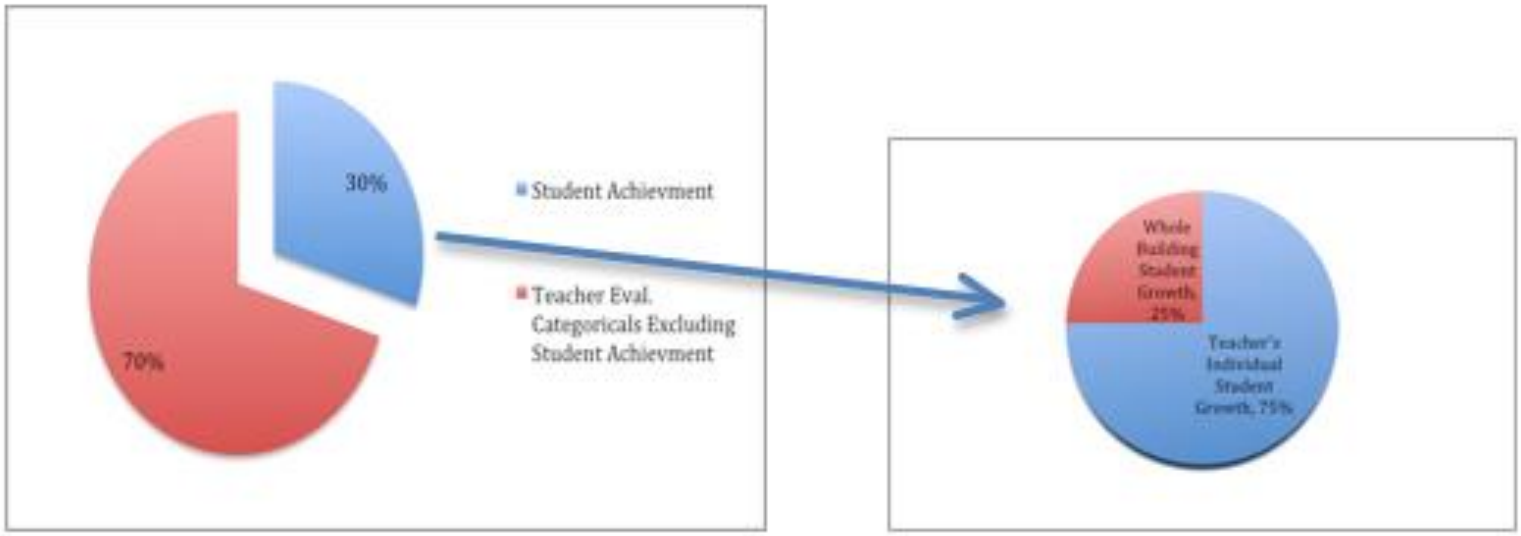
- Quantitative Measures
  - Measuring Student Growth
  - Represents 30% of the Total Teacher Evaluation

Ensuring Fair and Reliable Measures of Effective Teaching, Bill and Melinda Gates foundation

- Found that more effective teachers not only caused students to perform better on state tests, but they also caused students to score higher on other, more cognitively challenging assessments in math and English.
- Teaching is too complex for any single measure of performance to capture it accurately. Identifying great teachers requires multiple measures.

Teacher Evaluation Student Growth Measures

Figure 1. Thirty percent (30%) of teacher evaluation is based upon student growth and subdivided according to the figure below.



Which assessments should be used?

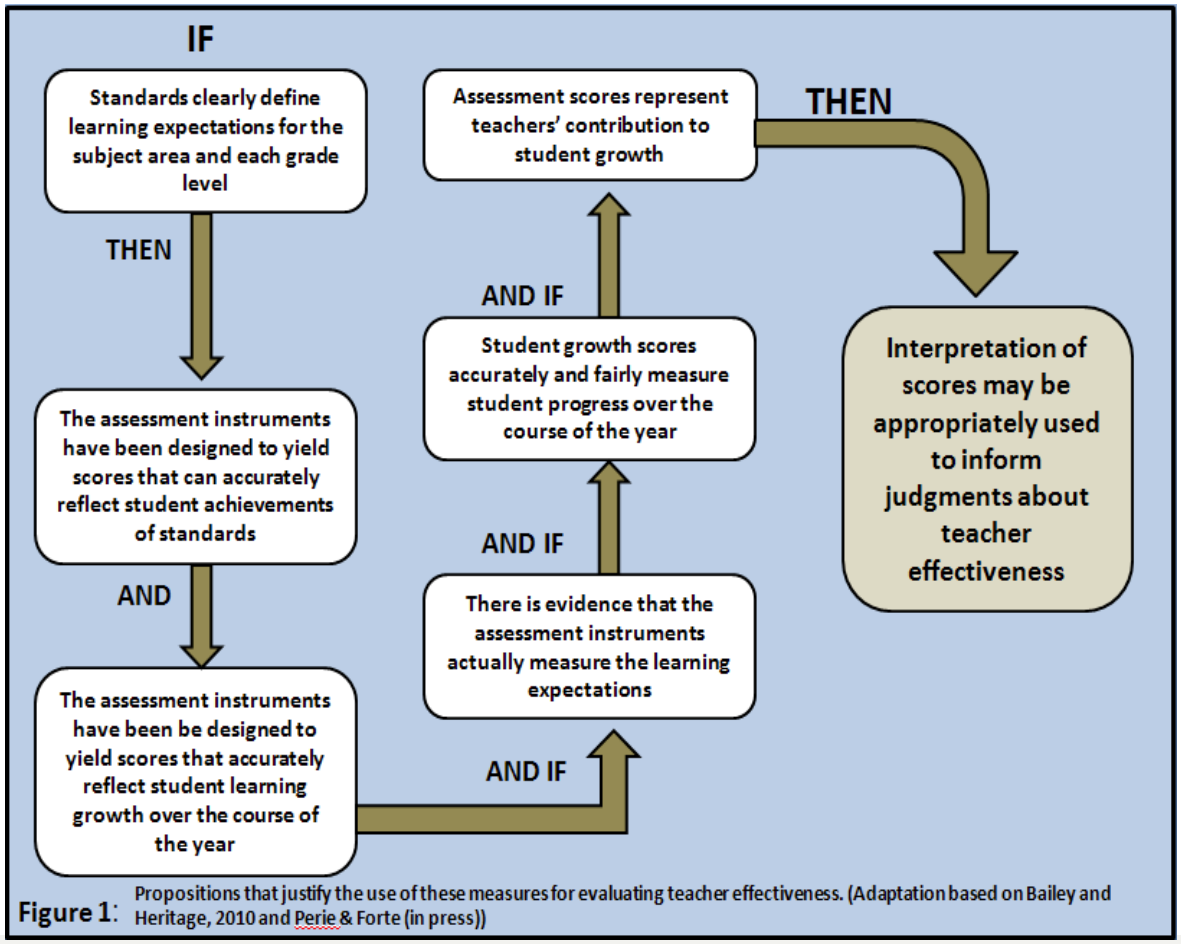


Figure 1: Propositions that justify the use of these measures for evaluating teacher effectiveness. (Adaptation based on Bailey and Heritage, 2010 and Perie & Forte (in press))

Teacher Evaluation Student Growth (K-6 Model)

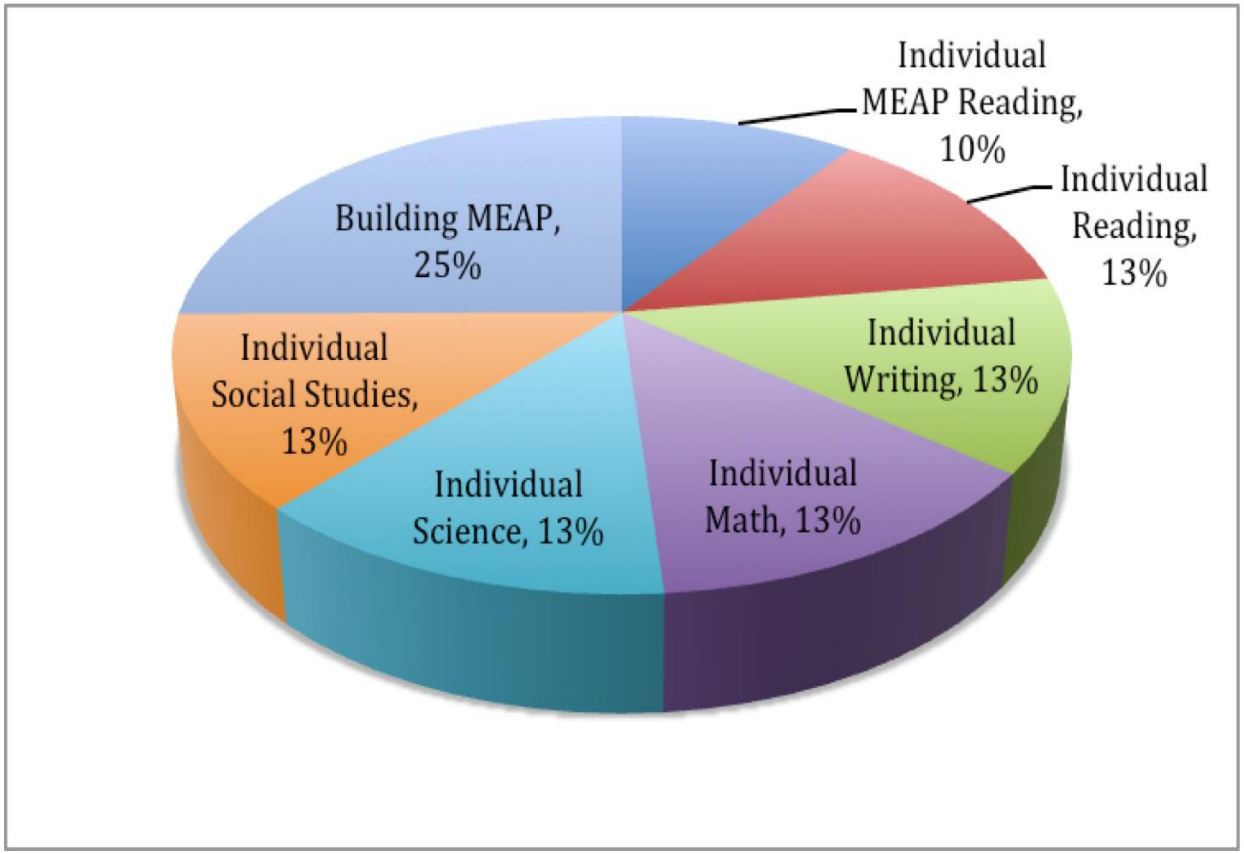
75% of each individual teacher's performance will be based on a combination of the following assessments.

Local Math Assessments (13%)
Local Reading Assessments (13%)
Local Writing Assessments (13%)
Local Science Assessment (13%)
Local Social Studies Assessment (13%)
Reading MEAP (10%)
*Elective Teacher (75%)

25% of each individual teacher's performance will be based on the entire building's MEAP scores.

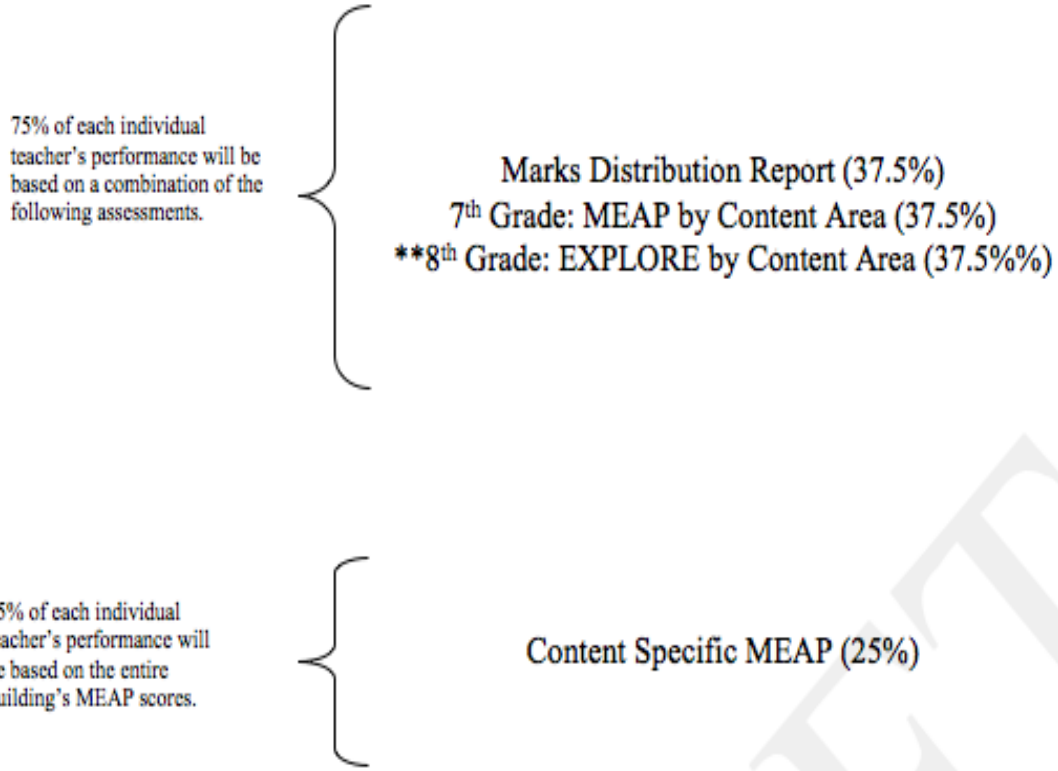
Reading MEAP (25%)

Elementary Student Growth (K-6 Model)



Teacher Evaluation Student Growth (7-8 Model)

Middle School (See figure 3):



- Variety in the Types of Assessment being utilized is key to reliability.
- Not all results should be based upon one measure e.g. MEAP or MME.

Teacher Evaluation Student Growth (9-12 Model)

High School (See figure 3):

75% of each individual teacher's performance will be based on a combination of the following assessments.

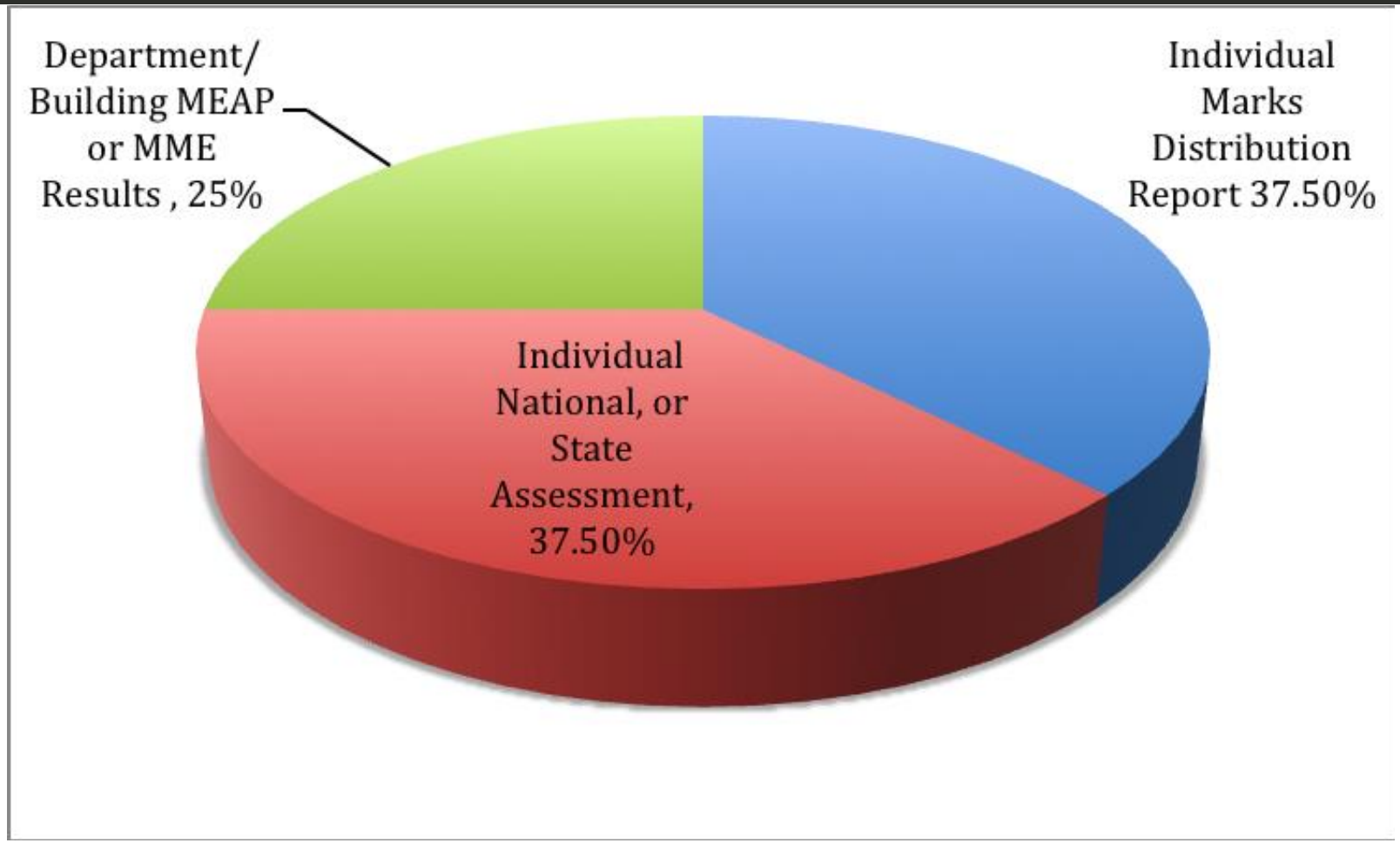
- All Grades—Marks Distribution Report (37.5%)
- **All Grades, Electives—Marks Distribution Report (75%)
- 9th Grade—EXPLORE to PLAN (37.5%)
- 10th Grade—PLAN to Practice ACT (37.5%)
- 11th Grade—Practice ACT to ACT (37.5%)
- 12th Grade—ACT to Practice ACT (37.5%)

25% of each individual teacher's performance will be based on the entire building's MEAP scores.

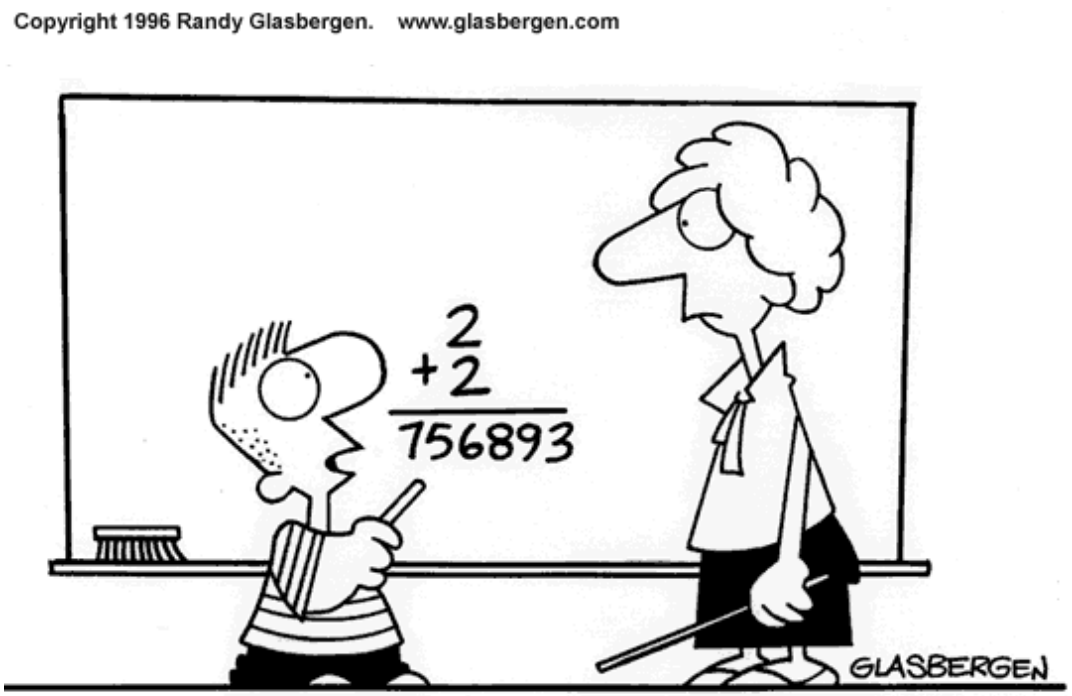
- MME (25%)

- Variety in the Types of Assessment being utilized is key to reliability.
- Not all results should be based upon one measure e.g. MEAP or MME.

Middle and High School Student Growth Model



Complexity and New Answers



"In an increasingly complex world, sometimes old questions require new answers."

Sample Rubrics

- K-6 Student Growth Effectiveness Rubric for Individual Teachers on Local and State Assessments (75%)
- 7-8 Student Growth Effectiveness Rubric for Marks Distribution Report or Locally Developed (Final Exam) Pre and Post Assessment (37.5%)
- 7-8 Student Growth Effectiveness Rubric for Individual Teachers on State Assessments (37.5%)
- 9-12 Student Growth Effectiveness Rubric for Individual Teachers on National Assessments (37.5%)
- 9-12 Student Growth Effectiveness Rubric for Building MME Assessment (25%)

K-6 Student Growth Effectiveness Rubric for Individual Teachers on Local and State Assessments (75%)

Highly Effective (4): At least 90% of students demonstrated adequate growth, as defined by the **district content area team**, on the end of year pre-post assessment.

For purposes of the individual MEAP score, 90% of students demonstrate **growth or proficiency** from one year to the next. **Growth is defined as scoring at least an equivalent score on the MEAP from one year to the next or demonstrating improvement from one category to another on the MEAP exam. For example a student that scores a 3M to a 3M when comparing 3rd and 4th grade MEAP scores demonstrated one year of growth. Proficiency is defined as a student that scores a 1 or a 2 on the MEAP.**

K-6 Student Growth Effectiveness Rubric for Individual Teachers on Local and State Assessments (75%)

Effective (3): Between **75% and 89%** of students demonstrated adequate growth, as defined by the district content area team, on the end of year pre-post assessment.

Minimally Effective (2): Between **51% and 74%** of students demonstrated adequate growth, as defined by the district content area team, on the end of year pre-post assessment.

Ineffective (1): Less than 51% of students demonstrated growth adequate growth, as defined by the district content area team, on the end of year pre-post assessment.

7-8 Student Growth Effectiveness Rubric for Marks Distribution Report or Locally Developed (Final Exam) Pre and Post Assessment (37.5%)

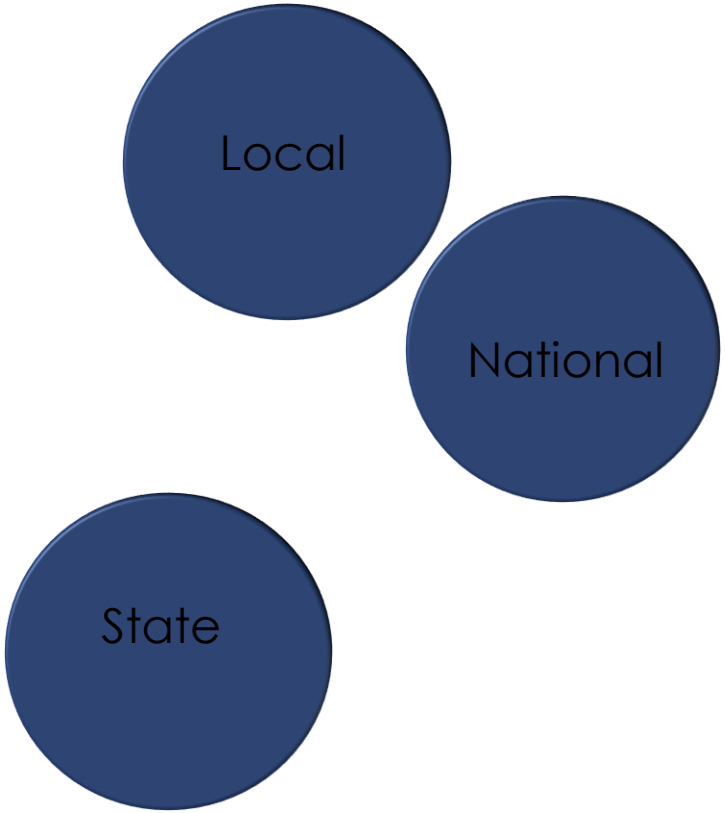
Highly Effective (4): The teacher has **two options** with respect to this measure: 1**) At least 90% of students show proficiency on the grade distribution report for final semester grades**. Proficiency will be considered a C average. 2) **The final exam will be given as a pre and post exam in order to determine student growth.** At least 90% of students demonstrated adequate growth, as defined by the district content area team, on the end of year pre-post assessment.

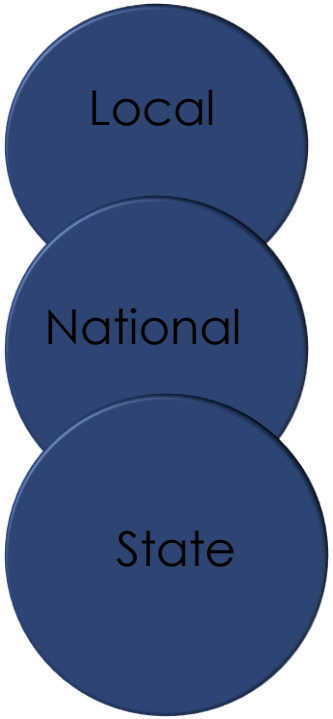
Effective (3): Convert the 90% to 75%-89%.

Minimally Effective (2): Convert 75%-89% to 51%-74%.

Ineffective (1): Convert 51%-74% to less than 51%.

Assessment Alignment





7-8 Student Growth Effectiveness Rubric for Individual Teachers on State Assessments (37.5%)

Highly Effective (4): For purposes of the individual MEAP score, 90% of students demonstrate growth or proficiency from one year to the next. Growth is defined as scoring at least an equivalent score on the MEAP from one year to the next or demonstrating improvement from one category to another on the MEAP exam. For example a student that scores a 3M to a 3M when comparing 7th and 8th grade MEAP scores demonstrated one year of growth. Proficiency is defined as a student that scores a 1 or a 2 on the MEAP.

Effective (3): Convert the 90% to 75%-89%.

Minimally Effective (2): Convert 75%-89% to 51%-74%.

Ineffective (1): Convert 51%-74% to less than 51%.

9-12 Student Growth Effectiveness Rubric for Individual Teachers on National Assessments (37.5%)

Highly Effective (4): At least 75% of students demonstrated growth or proficiency from one national assessment to the next. One year's growth on a national assessment e.g. EXPLORE to PLAN will be considered 1 point for one year of instruction. As for proficiency, the nationally established cut scores should be used based upon the assessment taken.

Effective (3): Convert the 75% to 50%-74%.

Minimally Effective (2): Convert 50%-74% to 25%-49%.

Ineffective (1): Convert 25%-49% to less than 25%.

9-12 Student Growth Effectiveness Rubric for Building MME Assessment (25%)

Highly Effective (4): The building or department **exceeded the growth criteria or state score on the MME**. Growth criteria will be based upon the target goal established in the building's school improvement plan. The minimum student growth threshold on the state assessment that is adopted by the building's school improvement team must align with the safe harbor calculation for making AYP. This calculation must be used for purposes of creating the objective statement in the school improvement plan.

- Safe Harbor Calculation: The Previous Year's % Proficient + 10% (100 - The Previous Year's % Proficient)

- For Example: 70% + [10% (100 – 70)] = 73%

9-12 Student Growth Effectiveness Rubric for Building MME Assessment (25%)

Effective (3): The building or department scored **within two (2) points of the growth criteria or state score on the MME**. Growth criteria will be based upon the target goal established in the building's school improvement plan. The minimum student growth threshold on the state assessment that is adopted by the building's school improvement team must align with the safe harbor calculation for making AYP. This calculation must be used for purposes of creating the objective statement in the school improvement plan.

- Safe Harbor Calculation: The Previous Year's % Proficient + 10% (100 - The Previous Year's % Proficient)

- For Example: 70% + [10% (100 – 70)] = 73%

9-12 Student Growth Effectiveness Rubric for Building MME Assessment (25%)

Minimally Effective (2): The building or department **scored at a level between the previous year's achievement score and within two (2) points** of the minimum safe harbor calculation. Growth criteria will be based upon the target goal established in the building's school improvement plan. The minimum student growth threshold on the state assessment that is adopted by the building's school improvement team must align with the safe harbor calculation for making AYP. This calculation must be used for purposes of creating the objective statement in the school improvement plan.

- Safe Harbor Calculation: The Previous Year's % Proficient + 10% (100 - The Previous Year's % Proficient)

- For Example: 70% + [10% (100 – 70)] = 73%

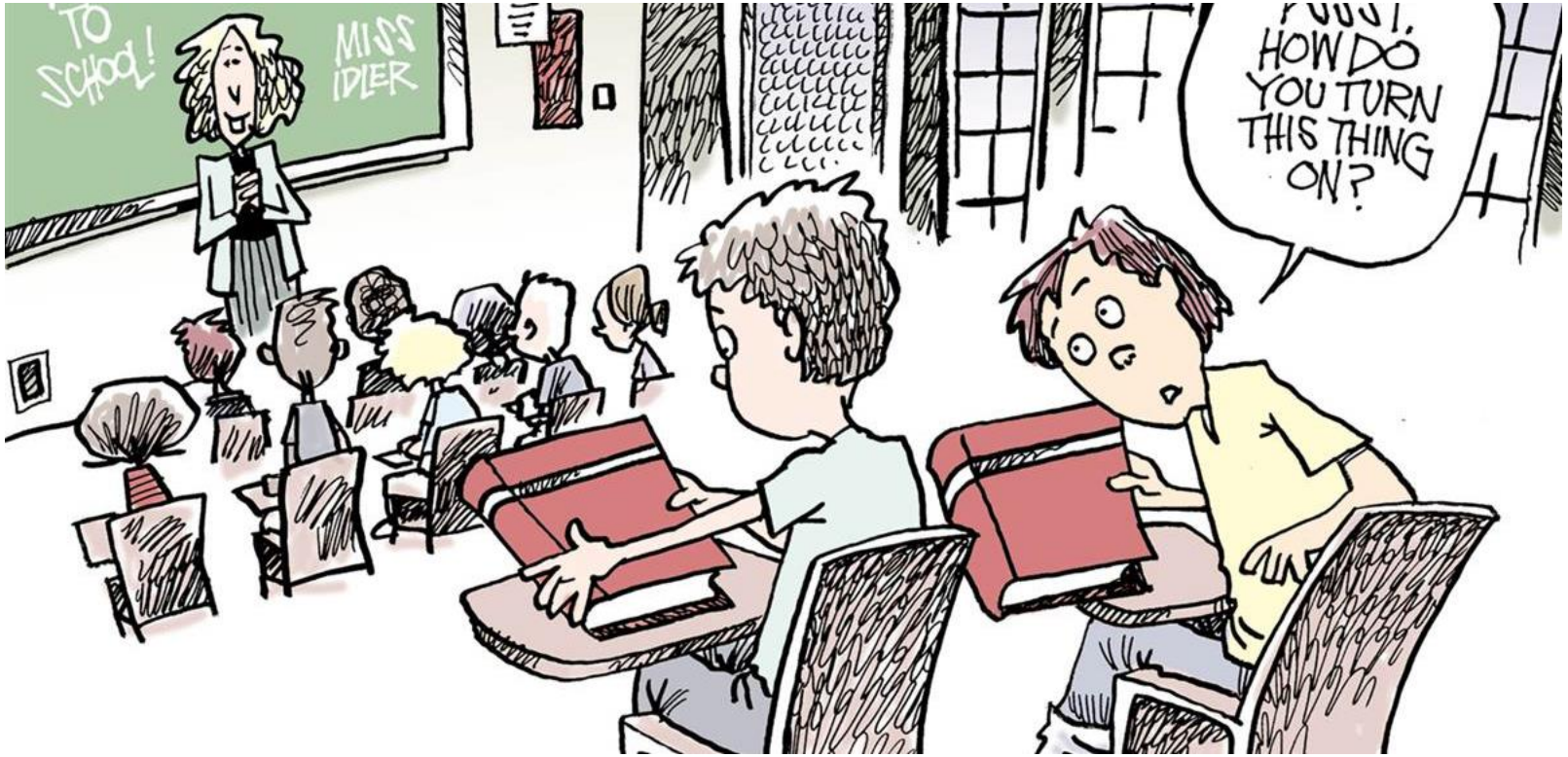
9-12 Student Growth Effectiveness Rubric for Building MME Assessment (25%)

Ineffective (1): The building or department **scored below the previous year's achievement score**. Growth criteria will be based upon the target goal established in the building's school improvement plan. The minimum student growth threshold on the state assessment that is adopted by the building's school improvement team must align with the safe harbor calculation for making AYP. This calculation must be used for purposes of creating the objective statement in the school improvement plan.

- Safe Harbor Calculation: The Previous Year's % Proficient + 10% (100 - The Previous Year's % Proficient)

- For Example: 70% + [10% (100 – 70)] = 73%

We Live in a Different Day and Age



Building Administrator Evaluation

Standards for Effective Administration

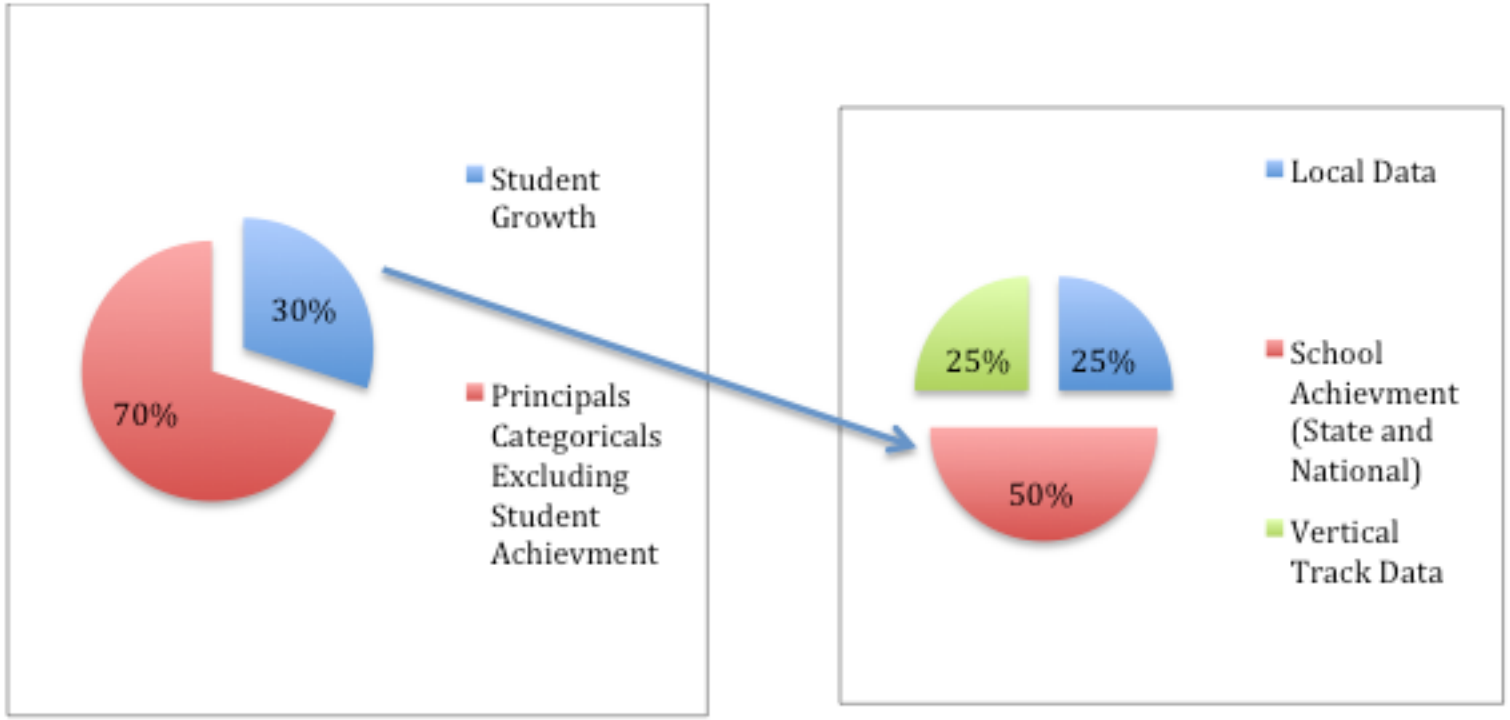
- Sources Adopted for Standards Development:

1. Interstate School Leaders Licensure Consortium's (ISLLC) Standards
2. AdvancED Standards for Quality Schools
3. Professional Learning Communities (PLC) Principles and Practices

Acknowledged by:
- National Policy Board for Educational Administration (NPBEA)
- American Association of School Administrators (AASA)
- National Association of Elementary School Principals (NAESP)
- National Association of Secondary School Principals (NASSP)
- Association for Supervision and Curriculum Development (ASCD)

Administrator Evaluation

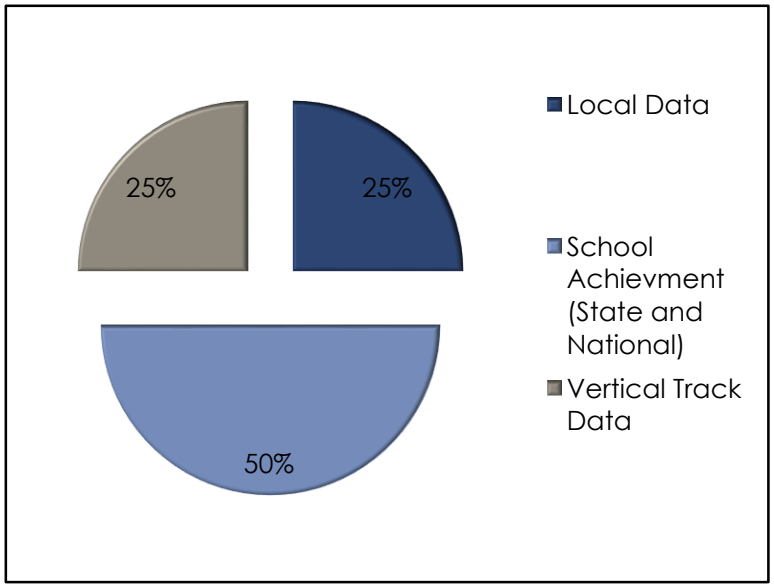
Figure 1. Thirty percent (30%) of administrator evaluation is based upon student growth



Building Administrator Professional Standards

Student Growth and School Achievement

- Local Data (25%)
- School Achievement (50%)
- Vertical Track Data (25%)



Local Data

- *Local Data (25%):* Twenty-five percent (25%) of the student growth portion will be based upon the collective building aggregate of the Marks Distribution Report (Secondary) or collective building aggregate of local assessments (Elementary).

School Achievement Data

- *School Achievement (50%):* Fifty percent (50%) of the student growth portion will be based upon the collective building aggregate on the State standardized assessment (MEAP/MME). These results will be determined based upon buildings reaching their safe harbor calculations or meeting state score averages.

Vertical Track Data

- *Vertical Track Data (25%):* Twenty-five percent (25%) of the student growth portion will be based upon State and/or National assessment results from multiple buildings. These results will be determined based upon buildings reaching their safe harbor calculations or meeting state score averages. At the K-6 level, a composite of all buildings will make up this portion of the administrator's evaluation. At the 7-12 level, assessment results will be based upon all buildings in the vertical track that a school belongs.

Administrator Evaluation Student Growth (K-6 Model)

Elementary (K-6, See figures 2, 3, 4 and 5):

25% of each administrator's performance will be based on a combination of the following assessments.

- Local Math Assessments (20%)
- Local Reading Assessments (20%)
- Local Writing Assessments (20%)
- Local Science Assessment (20%)
- Local Social Studies Assessment (20%)

50% of each administrator's performance will be based on the individual building's MEAP scores.

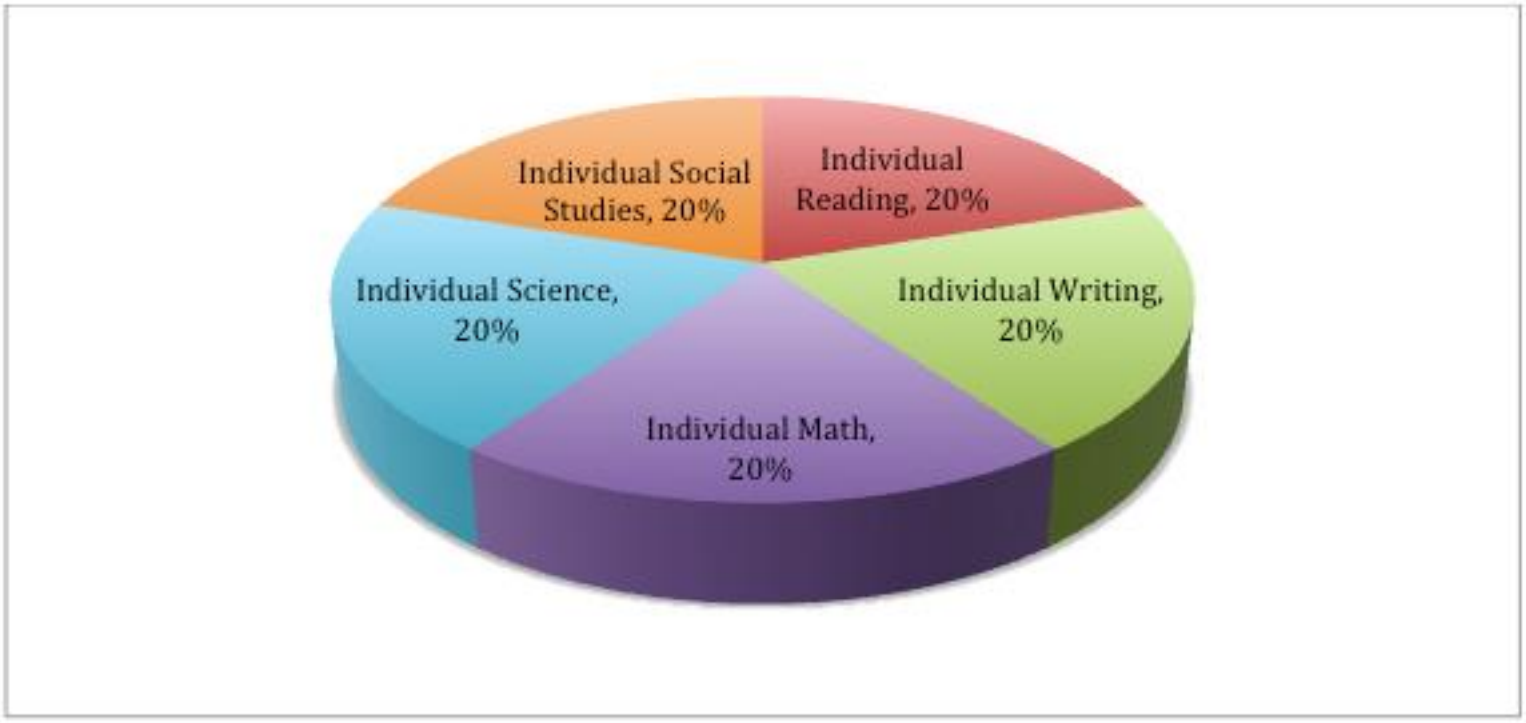
- MEAP, Composite of Tested Subjects (50%)

25% of each administrator's performance will be based on MEAP scores according to vertical track.

- Multi-Building Elementary MEAP (25%)

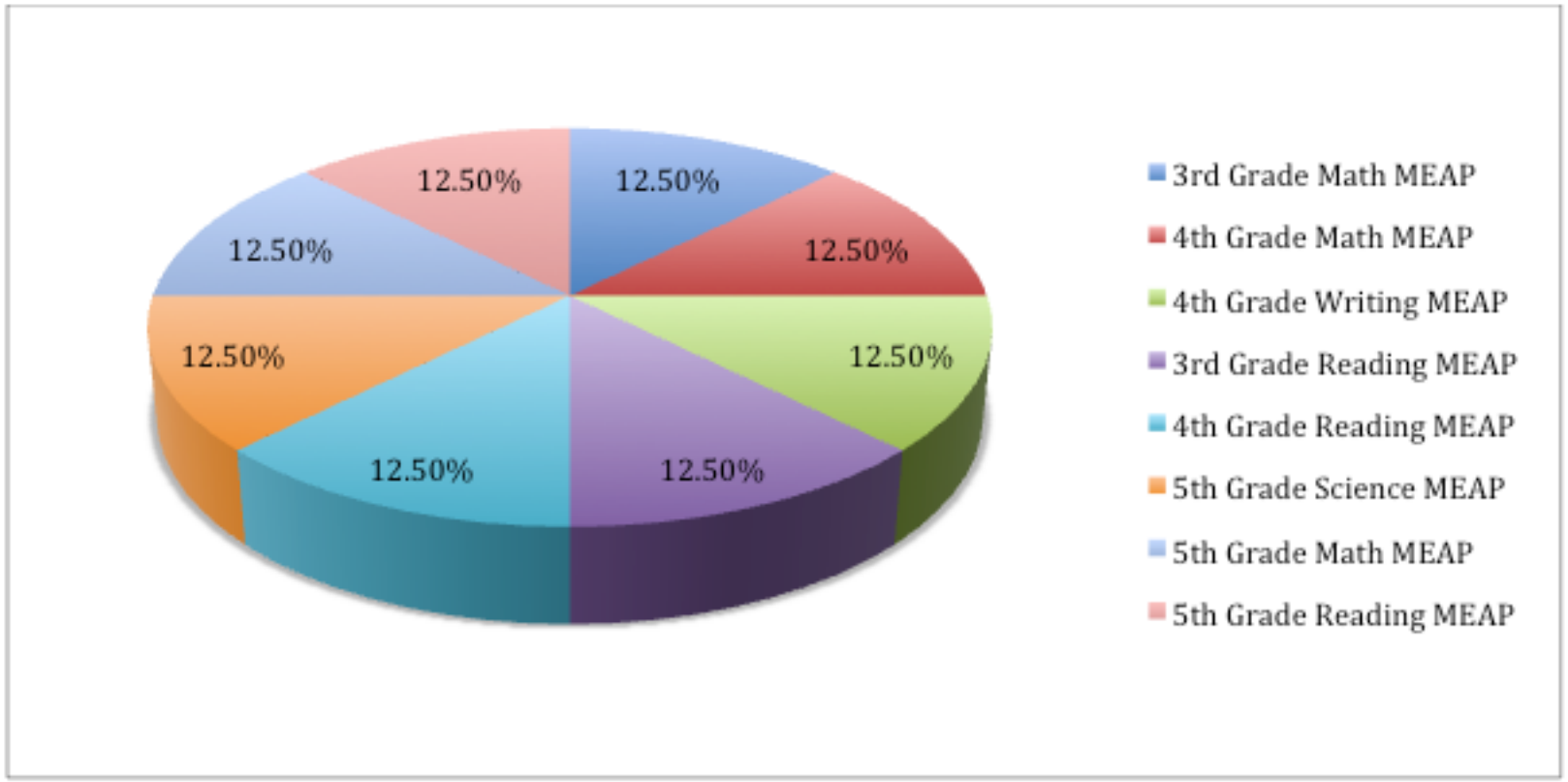
Administrator Evaluation Student Growth (K-6 Model)

Figure 2. K-6 student growth based upon Local Data measure (25%)



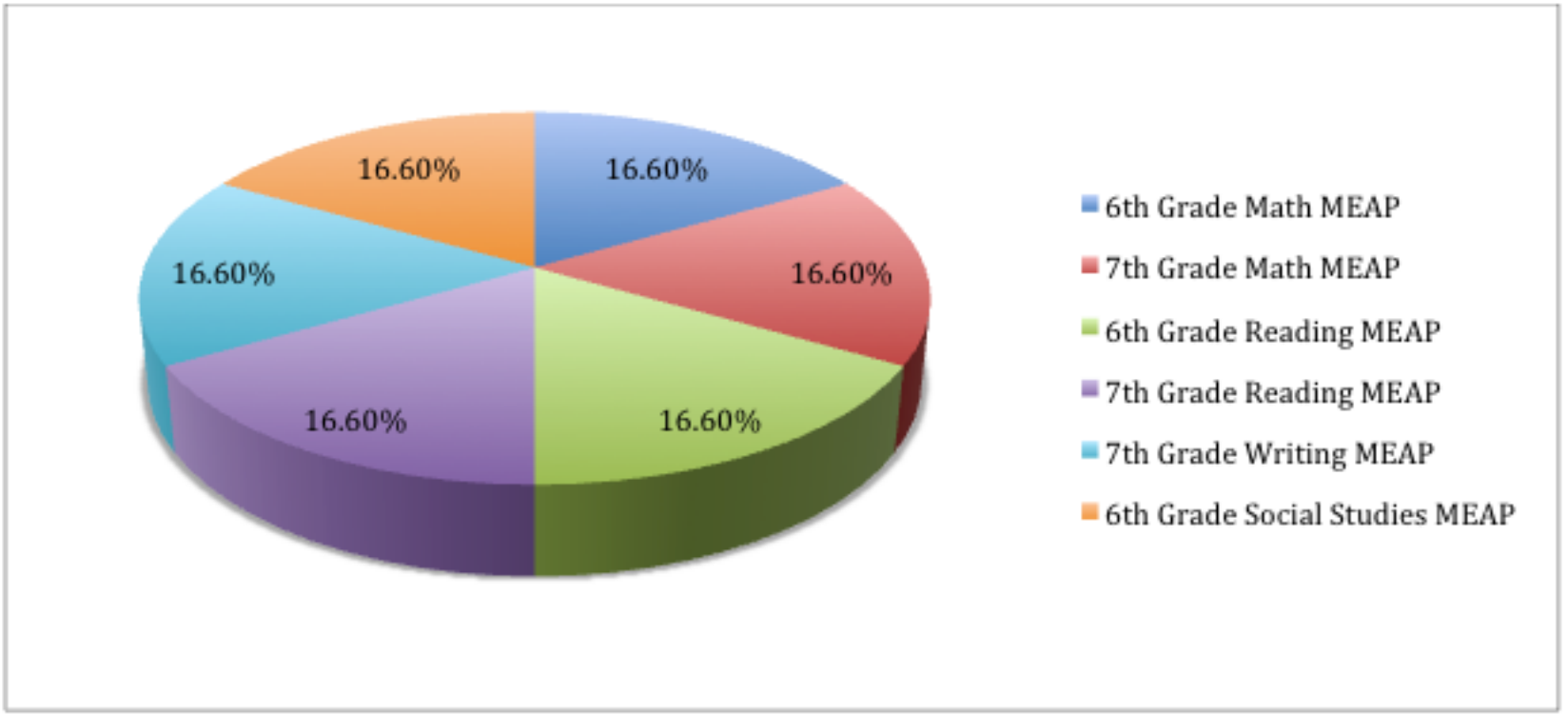
Lower Elementary Student Achievement Measure

Figure 3. Lower Elementary student growth based upon Student Achievement measure (50%)



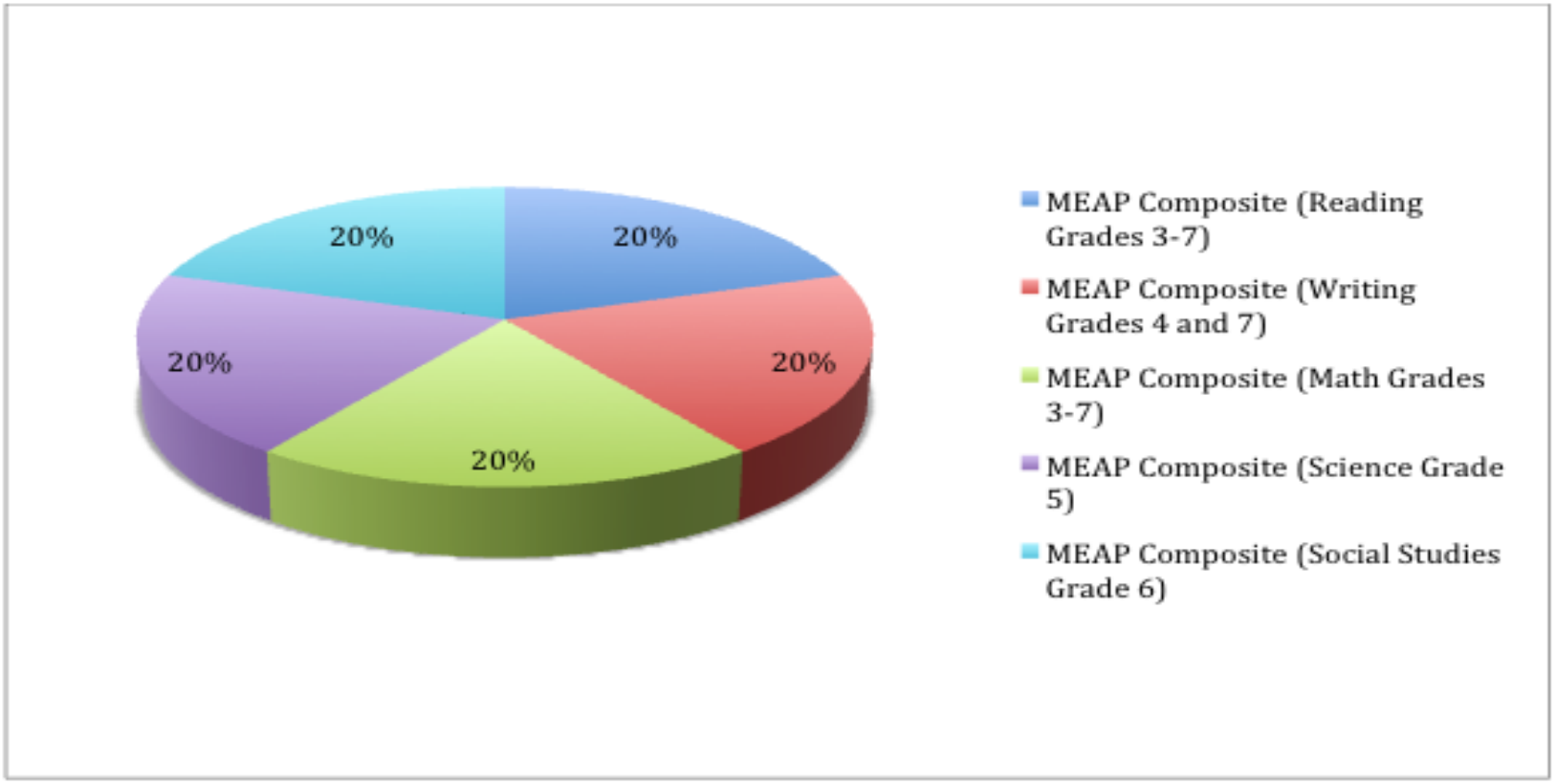
Upper Elementary Student Achievement Measure

Figure 4. Upper Elementary growth based upon Student Achievement measure (50%)



K-6 Vertical Track Measures

Figure 5. K-6 growth based upon Vertical Track measure (25%)



Administrator Evaluation Student Growth (7-8 Model)

Middle School (See figures 6, 7 and 8):

25% of each administrator's performance will be based on a combination of the following assessments.

- Marks Distribution Average (Math) (20%)
- Marks Distribution Average (English) (20%)
- Marks Distribution Average (Social Studies) (20%)
- Marks Distribution Average (Science) (20%)
- Marks Distribution Average (Electives) (20%)

50% of each administrator's performance will be based on the individual building's MEAP and National Assessment scores.

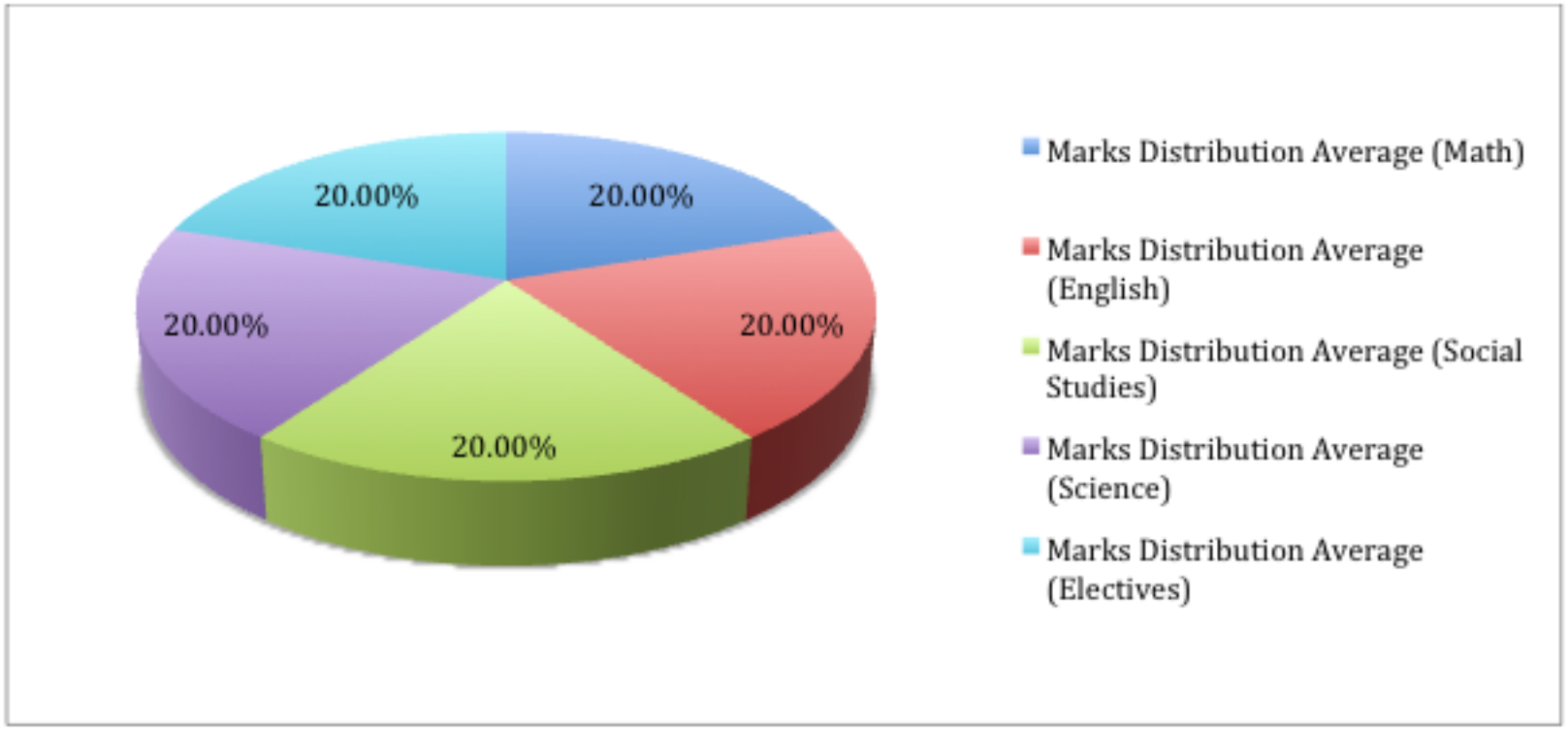
- MEAP and National Assessments, Composite of Tested Subjects (50%)

25% of each administrator's performance will be based on MEAP scores according to vertical track.

- Vertical Track MEAP/MME (25%)

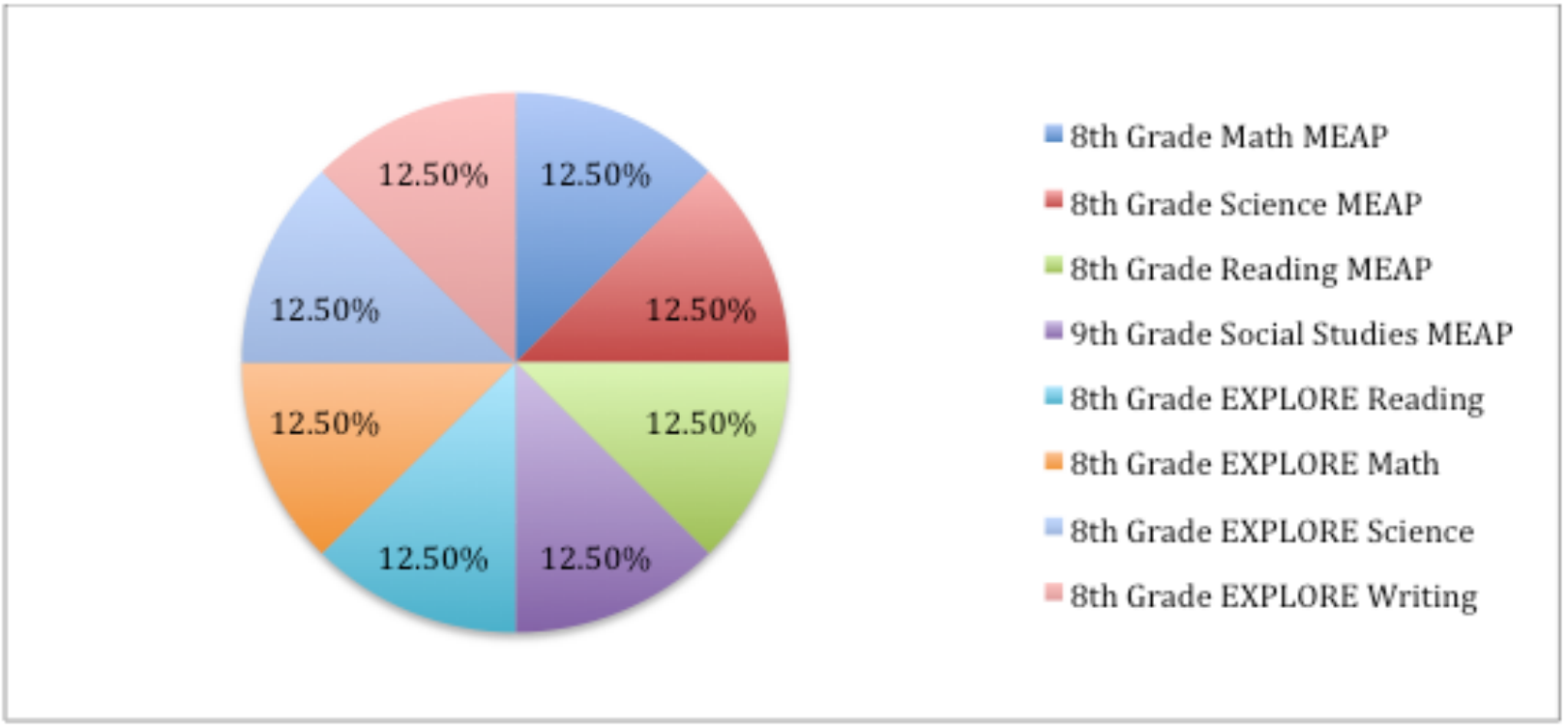
Middle and High School Local Data Measure

Figure 6. Middle and High School Local Data measure (25%)



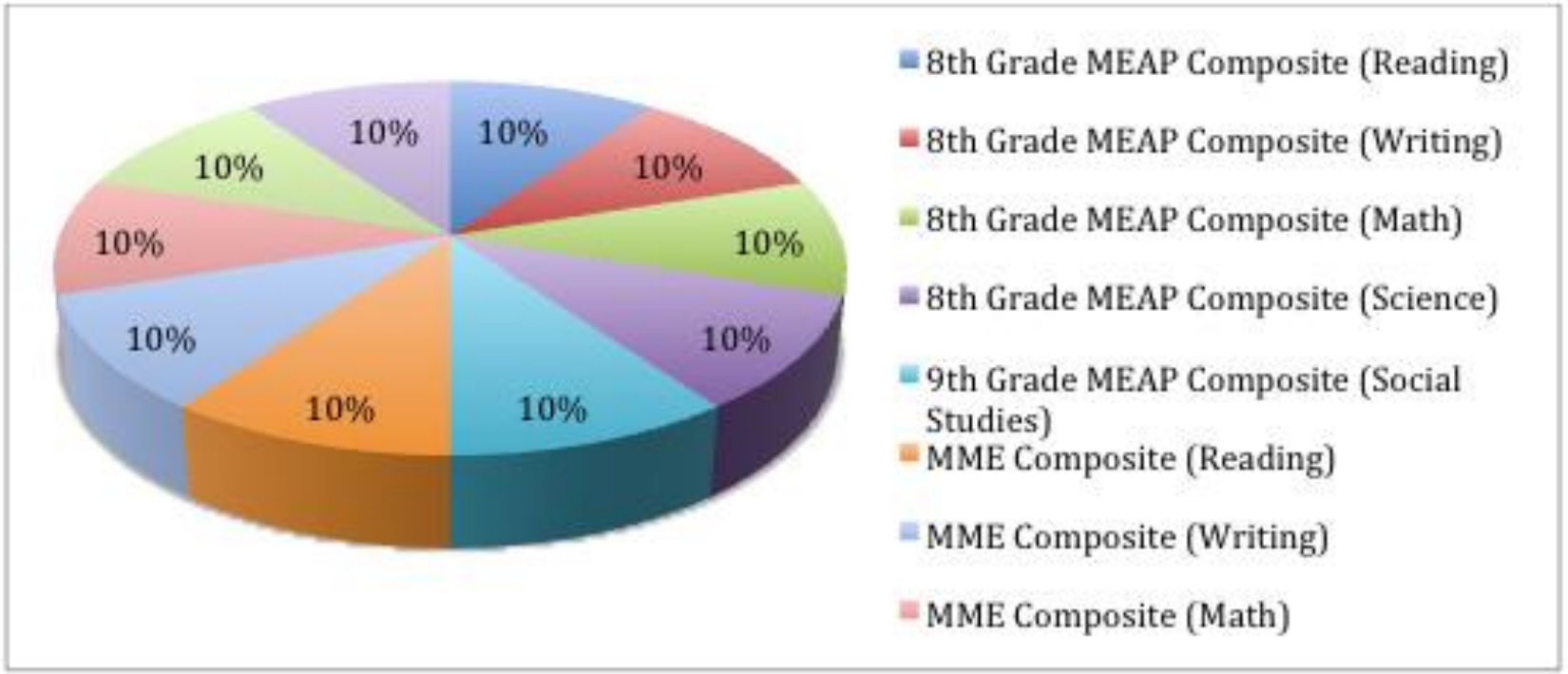
Middle School Student Achievement Measure

Figure 7. Middle School Student Achievement measure (50%)



Middle and High School Vertical Track Measure

Figure 8. Middle and High School Vertical Track Data measure (25%)



Administrator Evaluation Student Growth (9-12 Model)

High School (See figure 9, 10 and 11):

25% of each administrator's performance will be based on a combination of the following assessments.

- Marks Distribution Average (Math) (20%)
- Marks Distribution Average (English) (20%)
- Marks Distribution Average (Social Studies) (20%)
- Marks Distribution Average (Science) (20%)
- Marks Distribution Average (Electives) (20%)

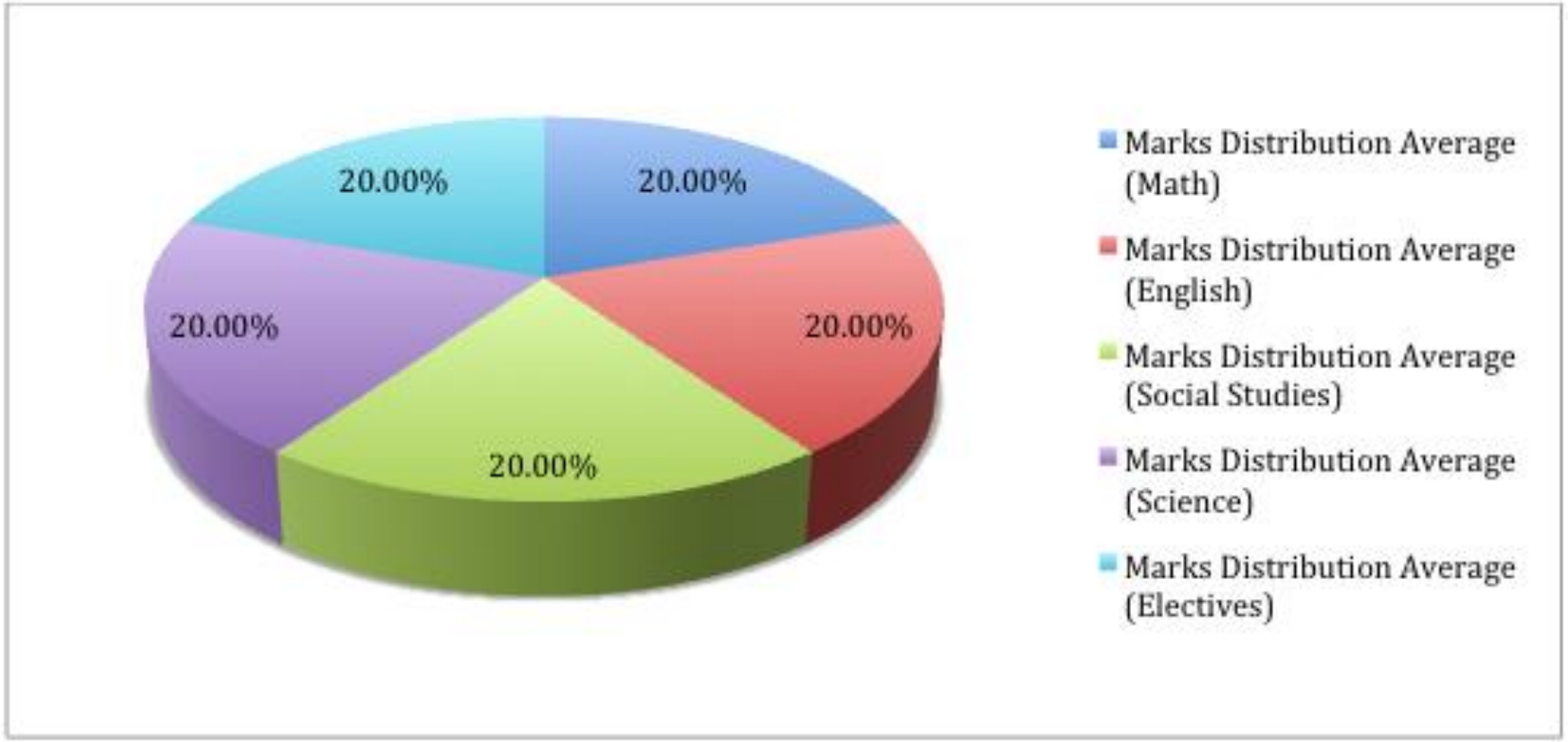
50% of each administrator's performance will be based on the individual building's MME/PLAN scores.

- MME/PLAN, Composite of Tested Subjects (50%)

25% of each individual administrator's performance will be based on MEAP scores according to vertical track.

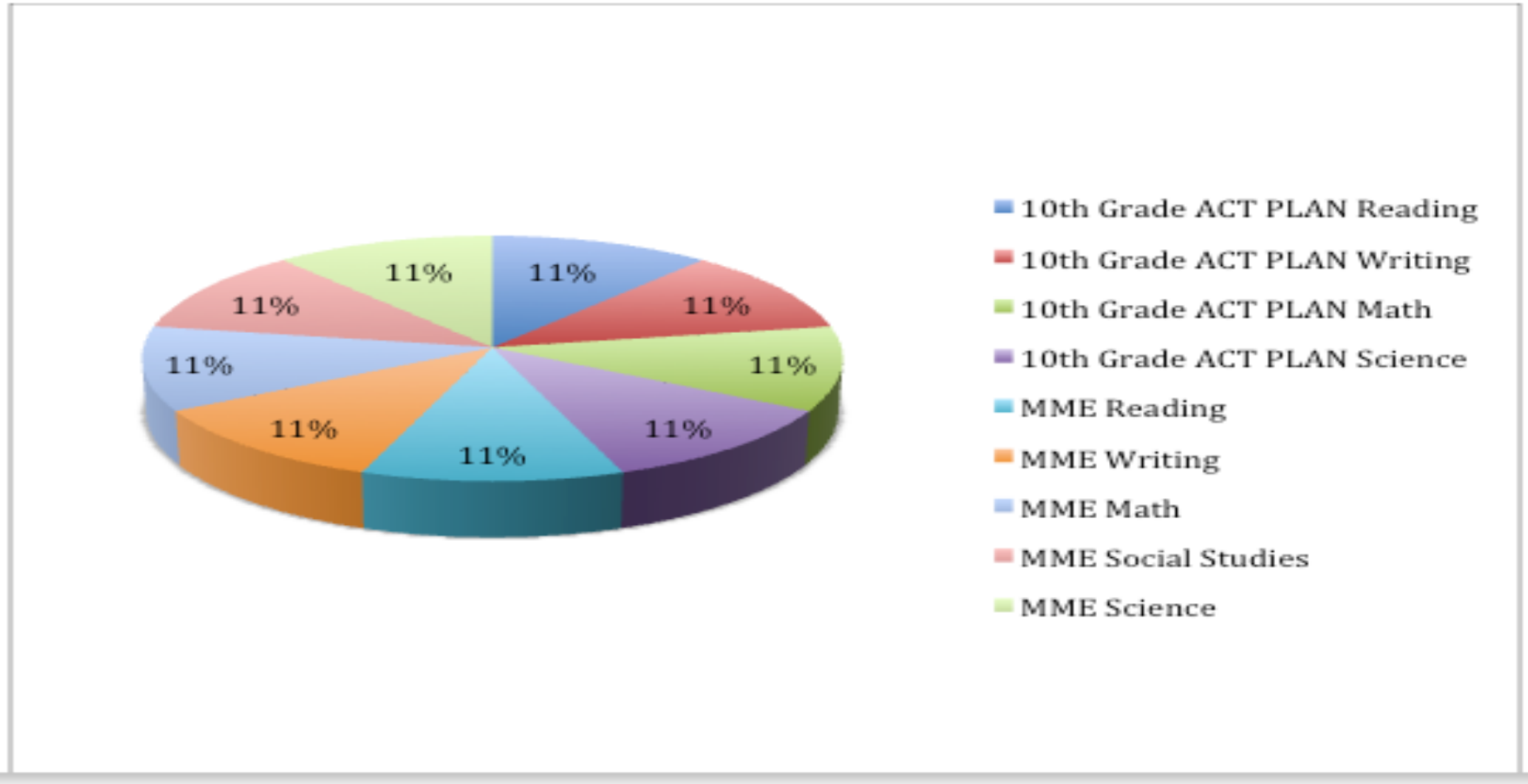
- Vertical Track MEAP/MME (25%)

Figure 9. Middle and High School Local Data measure (25%)



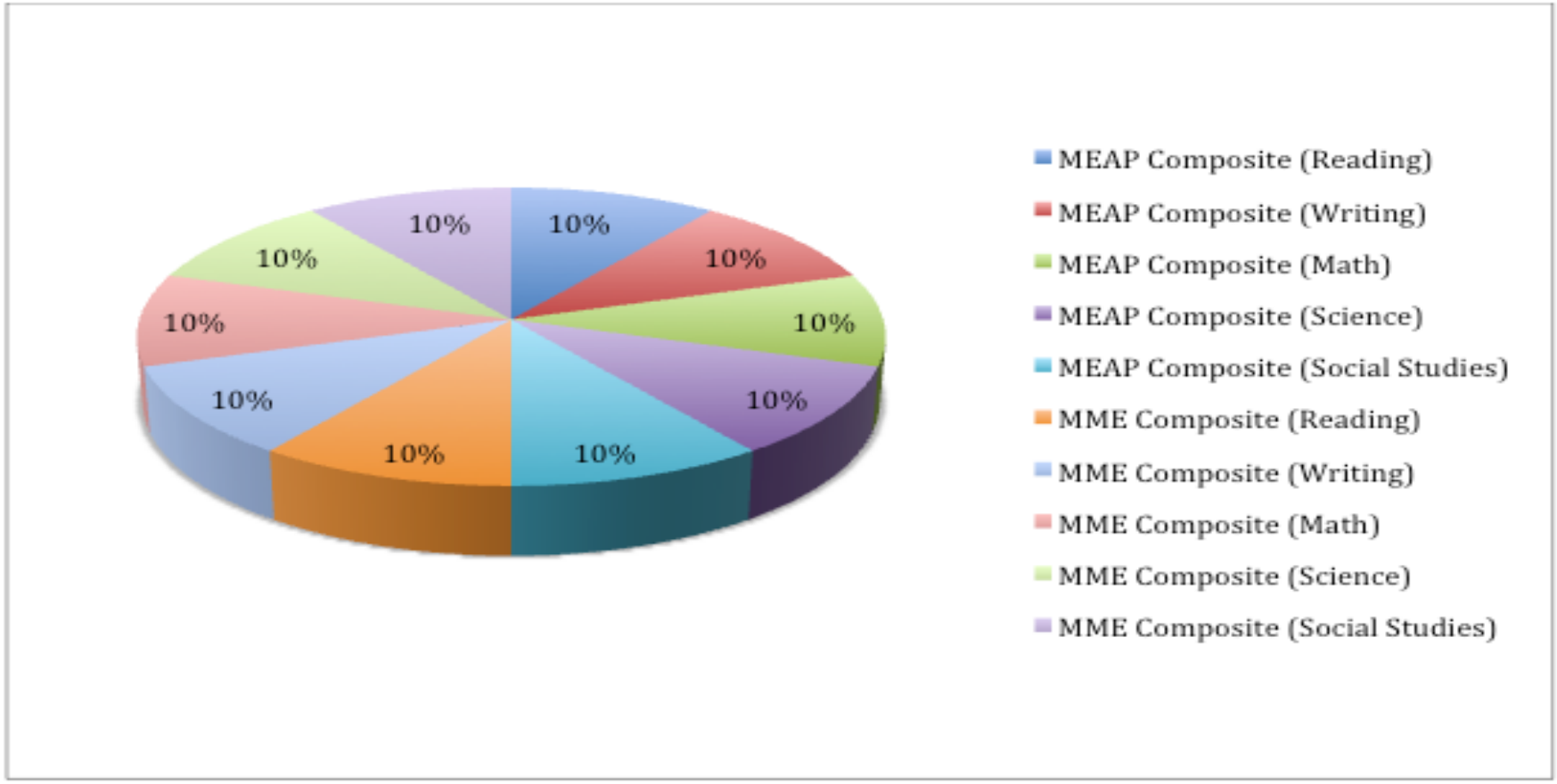
High School MME and PLAN Student Achievement Measure

Figure 10. High School MME and PLAN Student Achievement measure (50%)



Middle and High School Vertical Track Measure

Figure 11. Middle and High School Vertical Track Data measure (25%)



Administrator Rubrics

- Administrator rubrics were built upon the model established for teacher rubrics.
- Many similarities exist between the teacher evaluation model and the administrator evaluation model.

Results of Teacher Evaluation for Wayne-Westland Community Schools

- At the end of our first year we saw the following results:
  - 22.55% or 179 teachers rated as highly effective
  - 74.61% or 517 teachers rated as effective
    - **97.16% of teachers were reported as HE or E.**
  - 2.04% or 41 teachers rated as minimally effective
  - .8% or 1 teacher rated as ineffective
    - **2.84% of teachers were reported as ME or I**

Results of Teacher Evaluation for Wayne-Westland Community Schools

- At the end of our first year we saw the following results:
  - 23.68% or 179 teachers rated as highly effective
  - 68.39% or 517 teachers rated as effective
    - **92.07% of teachers were rated as HE or E**
  - 5.42% or 41 teachers rated as minimally effective
  - .13% or 1 teacher rated as ineffective
  - .13% or 1 teacher not completed
  - 2.25% or 17 teachers on leave
  - 25 teachers resigned rather than receive a minimally effective or ineffective rating which accounted for about 3% of Wayne-Westland teachers.
    - **10.93% of teachers were rated as ME, I, Not Complete or Resigned**

    - **NOTE: The 25 resignations are not calculated in the % above, but was added for purposes of understanding what will likely have transpired if they had not resigned.**

Keys to Success

- Communication is key!
  - "Wendy Thomson, the new chief executive of Neham Council…spent a full year conducting forums of discussion about a new vision, talking to thousands of people—only to find, from a survey, after 12 months that 80% of the people had never heard of the vision; and of the remaining 20%, 80% of them were against the vision? A year's hard work produces 4% in favor."—Fullan, M. (2010) *Motion Leadership, The Skinny on Becoming Change Savvy*.

Keys to Success

- Support of Wayne-County RESA related to the CLASS A Data Management System.
  - Data collection
  - Report collection
  - Met with the RESA team early on in the process
- Dedication of staff and support at the district level.
- Cooperation of the Teacher's Union.
- Consistent and ongoing discussion and training with principals
  - Helps create an understanding of what good teaching looks like

Must Do

- Must do
  - Develop a Frequently Asked Questions Document based upon questions from the staff
  - Ask for feedback and make appropriate adjustments
    - Every district is different
  - Provide for frequent and ongoing Professional Development
  - Vertical and Horizontal principal PLC meetings
    - Define evaluation, artifact, rubric interpretation
  - Constant reflection and revaluation
  - Give yourself permission to improve

Questions

Paul Salah Ed.D.

Deputy Superintendent for Educational Services

Wayne-Westland Community School District

Salahp@wwcs.mi.us.k12