Emerging Findings from the Measures of Effective Teaching Project (MET)

March 6, 2012
Why are we doing this presentation?
To share emerging findings from the Measures of Effective Teaching (MET) research project that is testing a variety of indicators for improving teaching and learning through better evaluation, feedback and professional development.

What are we seeking to gain from the presentation?
1. an understanding that combining observations, student surveys and achievement gains is a powerful way to identify practices that contribute to student growth and give teachers tailored, constructive insight for professional growth.

2. an understanding of how Pittsburgh’s measures to support Empowering Effective Teachers compares to emerging MET findings.
Overview of the Measures of Effective Teaching Project

✓ Unique, large scale research project funded by the Bill & Melinda Gates Foundation designed to identify valid and reliable indicators for better teacher evaluation, feedback and professional development

• 3,000 teachers in public schools across the country participating
• 7,500 lesson videos
• 900 trained observers
• 44,500 students completing surveys

✓ Next MET report expected mid-2012.
Three Key Take-Aways from MET

1. **High quality observations are critical**: Classroom observations that yield reliable results require clear standards, certified observers, and multiple observations per teacher.

2. **Multiple measures matter**: Combining three approaches (classroom observations, student feedback, and value-added) provides more useful information than any one approach.

3. **Combining new approaches to measuring effective teaching** – while not perfect – **significantly outperforms traditional measures**. Providing better evidence should lead to better decisions.
Take-away #1

High quality observations are critical
### Step 1: Define Expectations

Framework for Teaching (Danielson)

<table>
<thead>
<tr>
<th>Skill</th>
<th>Unsatisfactory</th>
<th>Basic</th>
<th>Proficient</th>
<th>Distinguished</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing student behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creating an environment of respect and rapport</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engaging students in learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managing classroom procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establishing a culture of learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using assessment in instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using questioning and discussion techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicating with students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Four Steps

- **Unsatisfactory**: Yes/no Questions, posed in rapid succession, teacher asks all questions, same few students participate.
- **Basic**: Some questions ask for student explanations, uneven attempts to engage all students.
- **Advanced**: Most questions ask for explanation, discussion develops/teacher steps aside, all students participate.
- **Distinguished**: All questions high quality, students initiate some questions, students engage other students.
Step 2: Ensure Accuracy of Observers

**TRAIN**
Observers learn to identify each competency at each performance level

**CERTIFY**
Observers rate pre-scored videos and must achieve adequate match with correct scores

**OBSERVE**
Observers are permitted to rate teacher practice

PASS

FAIL
Multiple Observations Result in Higher Reliability

Variation Due to Consistent Teaching Practice
Variation Due to Other Factors (Lesson, Rater, Section, Unexplained)

1 rater: 37%
2 raters: 53%
3 raters: 61%
4 raters: 66%

Each rater is observing a different lesson.

NOTES: The numbers inside each circle are estimates of the percentage of total variance in FFT observation scores attributable to consistent aspects of teachers' practice when one to four lessons were observed, each by a different observer. The total area of each circle represents the total variance in scores. These estimates are based on trained observers with no prior exposure to the teachers' students, watching digital videos. Reliabilities will differ in practice. See the research paper, Table 11, for reliabilities of other instruments.
Step 4: Verify Alignment with Outcomes

Teachers with Higher Observation Scores Had Students Who Learned More

State Math Tests

Teacher Value-Added Scores
(in months of schooling gained or lost)

Teacher Observation Scores (ranked by percentile)

Balanced Assessment of Mathematics

Teacher Value-Added Scores
(in months of schooling gained or lost)

Teacher Observation Scores (ranked by percentile)

State ELA Tests

Teacher Value-Added Scores
(in months of schooling gained or lost)

Teacher Observation Scores (ranked by percentile)

SAT9 Open-Ended ELA Test

Teacher Value-Added Scores
(in months of schooling gained or lost)

Teacher Observation Scores (ranked by percentile)

NOTES: Value-added estimated in student-level standard deviation units and converted to months of schooling using conversion factor of 0.25 standard deviations = 1 month of schooling. Slopes were calculated as linear regressions. Teachers’ value-added scores and observation scores from working with different groups of students.

The Pathway to the Promise.
What is Pittsburgh doing to ensure high quality observations?
Pittsburgh is addressing all four steps to ensure high quality observations

1. RISE defines expectations for teachers by identifying core competencies and clear definitions of proficiency.

2. Every observer is participating in the Instructional Quality Assurance and Certification Process of the Danielson Proficiency System to ensure accuracy and inter-rater reliability.

3. Multiple observations are conducted by multiple observers during a school year to ensure reliable results.

4. Pittsburgh is working with Mathematica Policy Research, Inc. to determine the alignment of our observations with outcomes.
Take-away #2
Multiple measures matter
TEACHING INDICATORS
from each teacher working with ONE GROUP of students:
- Classroom Observations
- Student Surveys
- Gains on State Tests
- Combination of Indicators

STUDENT OUTCOMES
from same teacher working with ANOTHER GROUP of students:
- Gains on State Tests
- Gains on Supplemental Tests
- Positive Student Feedback

Three Criteria:

Predictive power: Which measure could most accurately identify teachers likely to have large gains when working with another group of students?

Reliability: Which measures were most stable from section to section or year to year for a given teacher?

Potential for Diagnostic Insight: Which have the potential to help a teacher see areas of practice needing improvement? (We’ve not tested this yet.)
Measures have different strengths …and weaknesses

<table>
<thead>
<tr>
<th>Measure</th>
<th>Predictive power</th>
<th>Reliability</th>
<th>Potential for Diagnostic Insight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-added</td>
<td>H</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>Student survey</td>
<td>M</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>Observation</td>
<td>L</td>
<td>M/H</td>
<td>H</td>
</tr>
</tbody>
</table>
Is Pittsburgh using multiple measures?
Pittsburgh is Using Observations, Student Surveys and Value-Added to provide feedback for teaching.
✓ RISE observations are conducted district-wide.
✓ In 2011-12 all teachers of record have the opportunity to use student feedback collected through the district-wide administration of the Tripod Student Survey.

✓ Value-added and Tripod Student Surveys are being used to evaluate Career Ladder Teachers in 2011-12.

✓ The 2011-12 Career Ladders Teacher selection process is utilizing teacher value-added data in addition to classroom observations.

✓ For the first time in February 2012, approximately 40% of PPS teachers have teacher-level value-added data.

✓ 2010-11 Promise Readiness Cohort awards are based in part on team contributions to student growth (value-added).
Take-away #3

Combining new approaches to measuring effective teaching – while not perfect – significantly outperforms traditional measures.
Compared to MA Degrees and Years of Experience, the Combined Measure Identifies Larger Differences … on state tests

<table>
<thead>
<tr>
<th></th>
<th>State Math</th>
<th>State ELA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Master’s</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With</td>
<td>+.9</td>
<td>-4</td>
</tr>
<tr>
<td>Without</td>
<td>-.1</td>
<td>+2</td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 years or more</td>
<td>+2</td>
<td>+.3</td>
</tr>
<tr>
<td>3 years or less</td>
<td>-.3</td>
<td>-.4</td>
</tr>
<tr>
<td><strong>FFT, Student Survey, Value-Added on State Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 25%</td>
<td>+4.5</td>
<td>+1.2</td>
</tr>
<tr>
<td>Bottom 25%</td>
<td>-3.1</td>
<td>-1.3</td>
</tr>
</tbody>
</table>

*Months of Learning Gained or Lost*

**Average teacher**

*Notes: Value-added estimated in student-level standard deviation units and converted to months of schooling using conversion factor of 0.25 standard deviations = 9 months of schooling. Teachers’ value-added scores and scores of measures from working with different groups of students. Combined measure created with equal weights.*
Pittsburgh is on the right track and leading the way nationally to improving teaching and learning through better evaluation, feedback and professional development

- Pittsburgh’s multiple measures provide unprecedented feedback for teaching.
- Pittsburgh is serving as a model as state and federal policy focuses on effective teaching.
- Pittsburgh teachers and administrators are committed to working collaboratively.
- Multiple measures will be used for teacher evaluation in Pittsburgh starting in 2013-14 pending state legislation regarding teacher evaluation.