The Development of the STAR Classroom Observation Protocol and Results from 15,000 Classroom Observations
Introduction

• Objectives

– Provide an overview of the development and design of the STAR Protocol (a measure of classroom teaching and learning)

– To inform educators on the nature of teaching and learning occurring in classrooms and the extent classroom practices align with how students learn
STAR Protocol

What do you already know about the STAR Protocol?
# Getting Started

<table>
<thead>
<tr>
<th>What Do You Know?</th>
<th>What Do You Want to Know?</th>
<th>What Did You Learn?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Theory of Change

1. District-Level Practices & Environment
2. School-Level Practices & Environment
3. Classroom Practices & Environment
4. Student Learning
The History of the Reform Movement

• 1983 – 1988 “First Wave”
  – Top down, cosmetic change, first order change

• 1988 – 1993 “Second Wave”
  – Grass roots, performance based, second order change

• Since the Early 1990’s
  Perennialism  ➔ Essentialism
  Behaviorism  ➔ Cognitivism
  Teacher-centered  ➔ Student-centered
Why is this research important?

• Reform Context

  – 1993 House Bill 1209 passes establishing common learning goals for all students

  – Called for a fundamental change in what was taught, how it was taught, and how achievement was measured

  – Focus on alignment of curriculum and assessment and only recently on instruction (Baker, Clay, & Gratama, 2005)

  – Became standards-based: switched from the bell curve to the J-curve
Why is this research important?

- Standards-based teaching and learning takes into account the rapidly expanding field of brain-based research
  - the use of “research in neuroscience on how the brain works to gain an understanding of how students learn and develop in a classroom” (Madrazo and Motz, 2005, p. 56)
  - If the goal is to educate all students, then educators need to become knowledgeable about how the brain learns
  - Teacher role is changing; “the role of teachers is not to teach, per se, it is to ensure that students learn. Information can be presented to a student with great repetition and in astounding quantity, but only when that information passes into permanent memory does learning occur” (Bonnema, 2009, p. 2-3).

- Educators need support in determining whether the teaching and learning in their schools is aligned with how the brain learns and they need support in how to develop this type of teaching and learning
Why is this research important?

• **Post-secondary and workforce demands**
  – Sixty-seven percent of new jobs in the market today require some postsecondary education (Achieve Inc., 2006)
  – College students and new employees are being asked more and more to:
    • Engage in discussions
    • Make meaningful connections
    • Work with others
    • Create and deliver presentations
    • Write well-organized papers backed with evidence
    • Apply technology
    • Conduct research
  » (National Research Council, 2002)

• These tasks highlight the need for different instructional practices
• Reliable and valid instruments must be designed to accurately assess the extent to which classroom practices help prepare students for these types of tasks
# Instructional Changes Resulting from State Education Reform Acts 1988 - 1993

<table>
<thead>
<tr>
<th>Pre-reform</th>
<th>Post-reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-centered</td>
<td>Student-centered</td>
</tr>
<tr>
<td>Norm referenced</td>
<td>Criterion referenced</td>
</tr>
<tr>
<td>Bell curve</td>
<td>J curve</td>
</tr>
<tr>
<td>Teacher information</td>
<td>Student performance</td>
</tr>
<tr>
<td>Student compliance</td>
<td>Active inquiry</td>
</tr>
<tr>
<td>Adopted curriculum</td>
<td>Adapted curriculum</td>
</tr>
</tbody>
</table>
Instruction Input

Assessment Output

Curriculum Input
The Development of the STAR Classroom Observation Protocol
The STAR Instructional Framework

• 5 Essential Components
  – Skills
  – knowledge
  – Thinking
  – Application
  – Relationships
• 15 Indicators
• Multiple (95) Strategies
Development of the STAR Protocol

• Theoretical basis of the STAR stems from primary research examining effective teaching and learning, including the areas of Cognitive Science and theory, Instructional theory, and Learning theory.

• Development began in 2002 after reviewing this research and other existing reform-teaching observation protocols.

• The STAR Protocol really evolved out of the Teaching Attributes Observation Protocol (TAOP; Fouts, Brown, & Thieman, 2002), which is an instrument developed for use with school evaluation studies for the Gates Foundation and has a proven positive relationship to achievement.

• The TAOP has 27 indicators and contains constructs that are not vital to the measurement of teaching and learning aligned with brain-based research.
Design of the STAR Protocol

• STAR Protocol contains 15 Indicators organized around the four Washington State goals (Skills, Knowledge, Thinking, and Application)

• Also added a fifth area called Relationships to reflect the work of Newmann and Wehlage (1993) around supportive/collaborative learning environments

• Five Essential Components:
  - SKILLS: Did students actively read, write, or communicate?
  - KNOWLEDGE: Did students demonstrate depth of conceptual understanding?
  - THINKING: Did students demonstrate thinking through reflection and/or metacognition?
  - APPLICATION: Did students extend their learning into relevant contexts?
  - RELATIONSHIPS: Did interpersonal interactions reflect a supportive learning environment?
Design of the STAR Protocol

• Indicator and Strategy Development
  – Each of the five Essential Components is represented by a set of three Indicators.
  – Some Indicators directly from the TAOP, others from the Teacher Perspectives Questionnaire, others from a rubric created by the Northwest Regional Educational Laboratory, and from the work of Newmann and Wehlage
  – Indicators provide ways each Essential Component is manifested. The first Indicator in each set focuses on teacher methods; the following two Indicators focus on student cognitive processes and behaviors that demonstrate learning
  – Each Indicator includes multiple Strategies
  – The STAR contains approximately 100 examples of Strategies to manifest each Indicator in a lesson
  – Strategies were gathered from teacher reports and classroom observation data. This is not an exhaustive list, just a list of common observations.
What comments, questions, or reflections do you have about the development/design of the STAR Protocol?
Observer Training

- Observers participate in multiple trainings prior to conducting live observations. The goals of the training include:
  - 1) developing a common understanding and language around Powerful Teaching and Learning (read about development of protocol and research behind it)
  - 2) developing inter-rater reliability for using the instrument

- Training days consist of viewing videotaped lessons with other observers
- Each observer practices watching a 30-minute video while scanning and taking notes on the Protocol
- Each observer scores the Protocol individually
- Observers get in groups to review and discuss their scoring
- Observer pairs come to a consensus with their scores and present them to the entire group of observers
- Observers then shadow a more experienced researcher. The observer and researcher spend time after each observation to review scoring
Validity and Reliability of the STAR Protocol

• Validity
  – Established content validity by asking a group of twelve educators throughout the state of Washington to review the protocol
  – Feedback from reviewers was very positive regarding both the appropriateness of the items and the alignment of the protocol with reform goals
  – Established concurrent validity of the STAR by simultaneous scoring of 120 classrooms with the TAOP. Regardless of instrument the overall score remained consistent and the reliability between the two instruments was very high (Fouts, Brown & Thieman, 2002).
  – Established construct validity of the STAR by conducting Exploratory Factor Analysis. This revealed one dominant factor (we named it Powerful Teaching and Learning) that accounts for a large percentage of the variance in the scores.

• Reliability
  – The internal consistency estimate of reliability (Cronbach’s Alpha) for the 15 Indicators taken together is .92
Results from 15,000 Classroom Observations
The Sample

- Sample of convenience taken from a variety of program evaluation projects conducted over the last six years (2004 to 2010)

- Majority of schools in the sample are from Washington State; however a few schools from Hawaii, Missouri, and Oregon

- The sample of schools and classrooms in this study include more schools in ‘improvement’, more low-income schools, and more ethnically diverse schools than the state average for Washington State

- The STAR was designed to be used in a variety of subject area classrooms, but mainly observations are from core subject areas (language arts, mathematics, social studies, and science)
General Procedures

• Design of the STAR is unique in that it provides an aggregated measure of instructional practices in a school rather than an evaluation of a particular teacher.
• Therefore, sampling a larger number of classrooms is more important than spending longer periods of time in fewer classrooms.
• The observer’s duty is to record exactly what is seen and score the nature of the classroom activities in an objective way during the period he/she is in the classroom.
• The observation period ranges from 25-30 minutes in each classroom.
• Inter-rater reliability: every tenth classroom is observed by two researchers.
• During the 25-30 minute period all 15 Indicators are scored and an Overall score is given for each of the five Essential Components. An Overall score of 1-4 is then given to determine the presence of Powerful Teaching and Learning.
14% of the classrooms observed showed clearly observable evidence of Powerful Teaching and Learning (very aligned with our standards based reform movement).

Another 35% of classrooms observed demonstrated some evidence of PTL.

In the remaining 51% of classes, observers saw very little or no evidence of Powerful Teaching and Learning (we would consider not aligned with our standards based reform movement).
Why do we only see very clear evidence of PTL in 14% of the classrooms?
### Essential Component Frequencies

**Fall 2004 to Spring 2010 (n = 14,927)**

<table>
<thead>
<tr>
<th>Component</th>
<th>0=Not Observable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4=Clearly Observable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills</td>
<td>3%</td>
<td>9%</td>
<td>23%</td>
<td>40%</td>
<td>25%</td>
</tr>
<tr>
<td>Knowledge</td>
<td>5%</td>
<td>16%</td>
<td>33%</td>
<td>34%</td>
<td>12%</td>
</tr>
<tr>
<td>Thinking</td>
<td>8%</td>
<td>19%</td>
<td>32%</td>
<td>29%</td>
<td>11%</td>
</tr>
<tr>
<td>Application</td>
<td>28%</td>
<td>17%</td>
<td>23%</td>
<td>20%</td>
<td>11%</td>
</tr>
<tr>
<td>Relationships</td>
<td>16%</td>
<td>16%</td>
<td>52%</td>
<td>28%</td>
<td></td>
</tr>
</tbody>
</table>

- Which Essential Component scores the highest? Which scores the lowest?
### Indicator Frequencies

**Fall 2004 to Spring 2010 (n = 14,927)**

<table>
<thead>
<tr>
<th>Indicator</th>
<th>0=Not Observable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4=Clearly Observable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1</td>
<td>4%</td>
<td>8%</td>
<td>21%</td>
<td>39%</td>
<td>28%</td>
</tr>
<tr>
<td>Indicator 2</td>
<td>5%</td>
<td>16%</td>
<td>33%</td>
<td>34%</td>
<td>12%</td>
</tr>
<tr>
<td>Indicator 3</td>
<td>8%</td>
<td>19%</td>
<td>32%</td>
<td>29%</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 4</td>
<td>6%</td>
<td>12%</td>
<td>28%</td>
<td>34%</td>
<td>20%</td>
</tr>
<tr>
<td>Indicator 5</td>
<td>9%</td>
<td>17%</td>
<td>31%</td>
<td>30%</td>
<td>13%</td>
</tr>
<tr>
<td>Indicator 6</td>
<td>12%</td>
<td>20%</td>
<td>30%</td>
<td>27%</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Thinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 7</td>
<td>12%</td>
<td>16%</td>
<td>28%</td>
<td>27%</td>
<td>17%</td>
</tr>
<tr>
<td>Indicator 8</td>
<td>13%</td>
<td>18%</td>
<td>30%</td>
<td>28%</td>
<td>12%</td>
</tr>
<tr>
<td>Indicator 9</td>
<td>27%</td>
<td>23%</td>
<td>25%</td>
<td>18%</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 10</td>
<td>30%</td>
<td>16%</td>
<td>22%</td>
<td>20%</td>
<td>13%</td>
</tr>
<tr>
<td>Indicator 11</td>
<td></td>
<td>71%</td>
<td>7%</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>Indicator 12</td>
<td></td>
<td>92%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Relationships</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 13</td>
<td>12%</td>
<td>47%</td>
<td>39%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 14</td>
<td>19%</td>
<td>16%</td>
<td>23%</td>
<td>22%</td>
<td>20%</td>
</tr>
<tr>
<td>Indicator 15</td>
<td>14%</td>
<td>17%</td>
<td>27%</td>
<td>27%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Which Indicator scores the highest?

Which scores the lowest?
Overall Score Frequencies over 6-year period

- What do you conclude about change in PTL over the last 6 years?
What are some thoughts about how aligned instruction is with our standards-based reform?
STAR Scores by School Type

Comparing classrooms of different school levels, what patterns or differences do you notice?

Comparing these two tables, what patterns or differences do you notice?
### STAR Scores by Subject Matter

<table>
<thead>
<tr>
<th>Subject Matter</th>
<th>Skills</th>
<th>Knowledge</th>
<th>Thinking</th>
<th>Application</th>
<th>Relationships</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (n = 5349)</td>
<td>m 2.87</td>
<td>2.38</td>
<td>2.22</td>
<td>1.75</td>
<td>3.10</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>sd .94</td>
<td>1.01</td>
<td>1.09</td>
<td>1.34</td>
<td>.74</td>
<td>.91</td>
</tr>
<tr>
<td>Math (n = 4142)</td>
<td>m 2.79</td>
<td>2.34</td>
<td>2.24</td>
<td>1.21</td>
<td>3.02</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>sd .94</td>
<td>1.01</td>
<td>1.08</td>
<td>1.27</td>
<td>.81</td>
<td>.92</td>
</tr>
<tr>
<td>Science (n = 2000)</td>
<td>m 2.67</td>
<td>2.22</td>
<td>2.12</td>
<td>1.90</td>
<td>3.02</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td>sd 1.09</td>
<td>1.08</td>
<td>1.14</td>
<td>1.34</td>
<td>.84</td>
<td>.96</td>
</tr>
<tr>
<td>Social Studies (n = 1753)</td>
<td>m 2.61</td>
<td>2.11</td>
<td>2.03</td>
<td>2.07</td>
<td>2.93</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>sd 1.17</td>
<td>1.14</td>
<td>1.21</td>
<td>1.41</td>
<td>.84</td>
<td>.99</td>
</tr>
</tbody>
</table>

Comparing these subject areas, what patterns or differences do you notice?
PTL and Achievement

<table>
<thead>
<tr>
<th>Variables</th>
<th>PTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>.11&lt;sup&gt;t&lt;/sup&gt;</td>
</tr>
<tr>
<td>Math</td>
<td>.34*</td>
</tr>
<tr>
<td>Science</td>
<td>.19*</td>
</tr>
</tbody>
</table>

Correlations between Powerful Teaching and Learning and % of students passing different subject areas of the WASL.

<sup>t</sup><em>p < .06</em>  *<em>p < .01</em>  

As PTL goes up so does the % of students passing the WASL in these subject areas.
PTL and Achievement

• Follow-up regressions were performed using PTL to predict school level student achievement while controlling for low income

\[
\text{Achievement} = \text{Low-income} + \text{PTL}
\]

• Most notably, a unique contribution was found for PTL in predicting math achievement

• About 7% of the variance in math achievement was explained by PTL (statistically significant at the p < .001 level)
What comments, questions, or reflections do you have after looking at these findings?
Concluding Discussion Point

Is there a gap between the skills and knowledge taught in school today and what is needed to be successful in today’s economy?

If so, how must teaching and learning in classrooms change?