RESEARCH REPORT

BAKER ▪ EVALUATION ▪ RESEARCH ▪ CONSULTING

JULY 2010

STAR Protocol Report (Draft)

DUANE B. BAKER, Ed.D.
CANDACE A. GRATAMA, Ed.D.
KARI PETERSON, Ph.D.
CONNIE THOMPSON, Ph.D.

Duane Baker is the founder and president of Baker Evaluation, Research, and Consulting, Inc (*The* **BERC** *Group*). Dr. Baker has a broad spectrum of public school educational and program experience, including serving as a high school classroom teacher, high school assistant principal, middle school principal, executive director for curriculum and instruction, and assistant superintendent. In addition, he has served as an adjunct instructor in the School of Education at Seattle Pacific University since 1996, where his emphasis has been Educational Measurement and Evaluation and Classroom Assessment.

Dr. Baker also serves as the Director of Research for the Washington School Research Center at Seattle Pacific University. He also serves as an evaluator for several organizations including the Bill & Melinda Gates Foundation, Washington Education Foundation, Washington State Office of Superintendent of Public Instruction, and others.

Members of *The* **BERC** *Group* have K–20, experiences as teachers, counselors, psychologists, building administrators, district administrators, and college professors. The team is currently working on research and evaluation projects at the national, state, regional, district, school, classroom, and student levels in over 1000 schools in Washington State and nationally.

Table of Contents

Executive Summary

This study describes the development of the STAR Classroom Observation Protocol and presents results from classroom observations over the last six years. Research shows that the elements of current Washington State Education Reform efforts call for reliable and valid instruments that measure Powerful Teaching and Learning (PTL) in schools. As such, a conceptual framework was identified based largely on the foundational components of the Teaching Attributes Observation Protocol (TAOP), other similar measurements, and an extensive literature review. As a result, The BERC Group developed the STAR Protocol to measure Powerful Teaching and Learning in classrooms and schools. The protocol is divided into five Essential Components: Skills, Knowledge, Thinking, Application, and Relationships. Three Indicators are organized around each Essential Component, comprising a total of fifteen Indicators. Approximately 95 strategies are listed as part of the protocol and describe how the Indicators are manifested.

Following a training period, classroom observations were conducted in 14,927 classrooms. Provisions were made for regular inter-rater reliability and agreement checks where two raters scored the same observation and rated the protocol independently. Findings show a high degree of consistency in the rating process. Additionally, the Kappa reliability coefficient for the protocol is .90 suggesting good reliability between observers on the Overall score for the protocol. The results of a factor analysis show that the STAR Protocol is a uni-dimensional instrument tapping into a single construct: Powerful Teaching and Learning.

The general findings of this study are that strong Powerful Teaching and Learning was clearly observable in 14% of classrooms lessons. An additional 35% of classroom lessons show some evidence of Powerful Teaching and Learning. The remaining 51% show very little or no evidence of Powerful Teaching and Learning. Data also suggests little change has occurred in Powerful Teaching and Learning over a six-year period, when examining first-time schools that have never been previously observed. Although a moderate increase of 12% was found, this indicates the process for the reform movement is moving at a slow pace. The analyses also show a positive contribution of PTL to student achievement beyond the effects of low income. 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. Only small contributions of PTL were found for reading achievement (1%) and science achievement (2%). A small contribution of PTL was also revealed in predicting writing achievement (1%), however in the opposite direction.

The findings in this study show limited progress toward reform movement efforts. Research shows that the culture of the school must be fundamentally changed before academic achievement for students is likely. Meaningful reform can only move forward through Second Order changes. Creating a new school culture where professional development centers on researched-based instructional practices appears to be the key to providing the best chance for all students to be successful. To this end, the STAR Protocol was designed to synthesize characteristics of reformed instruction. This instrument offers a way to push the reform movement beyond its current state by being used as a measurement for reformed teaching and as a tool for professional development.

STAR Protocol Report

INTRODUCTION

The intent of this report is to provide a full description of the development of the STAR Protocol, to provide information on the reliability and validity of the Protocol, and to provide results from six years of classroom observation data. The results of the classroom observations will inform educators on the nature of teaching and learning occurring in classrooms and the extent classroom practices align with how students learn. The report will also highlight changes over time in classroom observation results using the STAR Protocol. Finally, the results of these observations will be used to determine whether Powerful Teaching and Learning can predict student achievement. This introductory section is followed by a description of the development of the STAR classroom observation protocol, design of the study, results, and summary and conclusions.

Why is this Research Important?

Subsequent to the 1980's and the publication of *A Nation at Risk*, states throughout the country adopted legislation to establish education standards for all students. In 1993 the passage of legislative House Bill 1209 was pivotal in supporting and transforming education in Washington State. This Education Reform Act defined Washington's reform movement, which established common learning goals for all Washington students. Educational goals were set with requirements that the academic experiences of students must be enriched with research-based curriculum, instruction, and assessment strategies. In other words, the effects of the legislation set a path to fundamentally change what was taught, the way it was taught, and the way to measure achievement. The forefront of most state, district, and school reform efforts to date have focused primarily on alignment of what to teach and what to test. Ideally, alignment of *how* to teach differently would have been developed simultaneously with *what* to teach and *what* to test, but it is only of late that the focus on aligned instruction has become a widespread interest (Baker, Clay, & Gratama, 2005).

Many are concerned about the emphasis placed on curriculum and assessment and the output in test scores provided by the current educational paradigm. The task before educators is clear in that all students must reach high levels of achievement regardless of ethnicity, poverty, or other disadvantaging circumstances. Although these instructional elements are more visible in the current reform efforts, there is fear that the overdue focus on effective, standards-based instructional practices will contribute to the delay of reaching this goal. But what did leaders expect from over a decade of *curriculum* reform and *testing*? It could be argued that much more has been gained from the system than should have been expected, given that one of the critical systems components was largely ignored (*instruction*). The current reforms seem to be headed in the right direction, but the past does suggest the need for an ongoing check on treatment fidelity. That is, a check to make sure both the curriculum and instructional components are supported systemically and effectively.

In spite of the recent attention focused on the alignment between what students learn in high school and what they need to know in college or the workforce, there appears to be a wide gap (Kuh, 2007). Sixty-seven percent of new jobs in the market today require some postsecondary education (Achieve Inc., 2006). New students face high expectations of post-secondary education such as, engaging in class discussions and making meaningful applications, working with others in and out of class, creating and delivering presentations, writing well-organized papers backed with evidence, applying technology to problem-solving, and developing well-designed lab projects. They are expected to analyze text, to draw inferences and conclusions, to conduct research, and to offer explanations to various phenomena (National Research Council, 2002). These tasks are fundamentally different than high school tasks and highlight the need for not only more rigorous course-taking in K-12 systems and aligned assessment from high school to college entrance, but high-impact instructional practices at the K-12 levels.

Educational researchers face the challenge of identifying clear constructs for teaching and learning that will provide a clear focus for examining the alignment of curriculum, instruction, and assessment. Instructional strategies aligned with reform content and assessments are described in various ways. Some refer to brain-based instruction, constructivist teaching, and reform-like teaching to describe classrooms characterized by active inquiry, in-depth learning, and performance assessment (Fouts, Brown & Thieman, 2002; National Research Council, 1999a; National Research Council, 1999b). In this report, we use the term *Powerful Teaching and Learning* (PTL) to encompass the elements and ideals of aligned and reformed instructional practices (Baker, 1998).

Comprehensive reform must have as its objective a change in standards (curriculum), testing (assessments), ***and*** pedagogy (teaching and learning). All three components are needed to fundamentally change schools. With the goal of helping all students regardless of race, gender, socioeconomic status, or learning disabilities reach proficient levels in academic areas, the need for high quality instructional practices are essential and must be seriously considered as one of the critical components along with aligning curriculum and assessment with the standards. As such, reliable and valid instruments must be designed to accurately assess the extent to which classroom practices are aligned with Powerful Teaching and Learning. Constructing classroom data observation instruments to measure the extent to which instruction is aligned with reform principles represents a shift from ignoring this critical element to pushing it to the forefront of the reform movement.

Over the last several years, members of The BERC Group have conducted thousands of classroom observations to determine the extent to which classroom instruction is aligned with brain-based research and reform efforts. PTL builds conceptual skills and knowledge, shows evidence of metacognition and personal reflection, is relevant to learners, and is supported by strong relationships. It embodies both teacher and student empowerment in a student-centered learning environment. It involves changes in: depth of content, roles of teachers and students, classroom interactions, and student understandings in problem solving and inquiry processes (Tittle & Pape, 1996).

Research Questions

The three research questions that guided our study were as follows:

1. To what extent are classroom practices aligned with how students learn?
2. To what extent have classroom practices changed fundamentally in the last six years?
3. Can Powerful Teaching and Learning predict student achievement?

DEVELOPMENT OF THE STAR CLASSROOM OBSERVATION PROTOCOL

The BERC Group began development of a classroom observation protocol in 2002 after extensive review of research from cognitive science, learning theory, and instructional theory. Additionally, researchers analyzed a variety of existing reform-teaching observation protocols (see list below). After reviewing the literature and existing measures, BERC Group researchers decided to use the Teaching Attributes Observation Protocol (TAOP; Fouts, Brown, & Thieman, 2002) as the basis for creating their classroom observation protocol. The TAOP was originally a research instrument designed for evaluations specific to the Gates Foundation. As such, the instrument was developed with additional constructs that were not vital to the measurement of teaching and learning aligned with brain-based research, thus in the creation of the new protocol all extraneous constructs were eliminated. These included indicators on classroom technology and assessment. In addition, the 27 indicators of the TAOP proved to be unwieldy for use during short classroom observations.

- Reformed Teaching Observation Protocol (Piburn et al., 2000)
- Sheltered Instruction Observation Protocol (Echevarria, Vogt, & Short, 2004)
- Standards of Authentic Instruction (Newmann & Wehlage, 1993)
- The Implementation of Alternative Assessment Procedures and Washington State Reform (Baker, 1998)
- Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement (Marzono, Pickering, & Pollock, 2001)
- The Art and Science of Professional Teaching: A Developmental Model for Demonstrating Positive Impact on Student Achievement (Simpson, 2001)
- Creating 21st Century Teachers for Washington State: Linking Professional Growth to Improved Student Learning (Simpson, 2001)

In order to ensure data collected from the protocol would be useful to schools, researchers decided to organize the new protocol around the four Washington State goals (Skills, Knowledge, Thinking, and Application)[1]. Researchers also added a fifth area called Relationships to reflect the

[1]Goal One: (Skills) Read with comprehension, write with skill, and communicate effectively and responsibly in a variety of ways;
Goal Two: (Knowledge) Know and apply the core concepts and principals of mathematics, social, physical, and life sciences; civics and history; geography; arts; and health and fitness;
Goal Three: (Thinking) Think analytically, logically, and creatively, and integrate experience and knowledge to form reasoned judgments and solve problems; and;
Goal Four: (Application) Understand the importance of work and how performance, effort and decisions directly affect career and educational opportunities.

work of Newmann and Wehlage (1993) around supportive/collaborative learning environments. Researchers decided on the acronym, STAR, for the new observation instrument. This process produced a protocol with five Essential Components and fifteen Indicators measuring the five Essential Components, with three Indicators falling under each component. The following sections contain the theoretical basis behind the Protocol, information on its development and design, as well as scoring procedures and results from about 15,000 observations using the Protocol.

Conceptual Framework of the STAR Protocol

The conceptual framework of the STAR Protocol consists of a constructivist or student-centered approach and is designed to detect the presence of Powerful Teaching and Learning in a classroom, a school or a district. There is a considerable amount of research that supports this idea, and the research has direct implications for how children should best be taught. The research stems from a depth of primary research examining effective teaching and learning, generated over the last fifteen years of standards-based, criterion-referenced, educational reform. In this new reform context, Cognitive Science and theory, Instructional theory, and Learning theory provide a foundation for Powerful Teaching and Learning to maximize student learning.

Cognitive Science

The books *How People Learn: Brain, Mind, Experience, and School* (National Research Council, 1999a) and *How People Learn: Bridging Research and Practice* (National Research Council, 1999b) provide a synthesis of the research behind Powerful Teaching and Learning. Extensive examination of developments in cognitive science over the last several decades indicates that students benefit most from instruction that manifests three constructs: active-inquiry, in-depth learning, and performance assessment. When teachers reported and were observed employing classroom practices aligned with these constructs, achievement improved for students. The five Essential Components of Powerful Teaching and Learning are at the heart of standards-based instruction – everyone in the building teaching in a way that helps every student meet the standards that drive contemporary educational reform efforts.

Cognitive Theory

Cognition mirrors information-processing systems. The mind, like a computer, stores information in neural networks that are activated and connected by thought processes. Teaching and learning aligned with cognitive theory effectively integrate new experiences with existing networks of knowledge. The more connections there are between networks, the deeper and more lasting learning becomes. This means that educators must facilitate experiences that allow students to discover new learning built on each student's existing knowledge base.

Instructional Theory

The switch (in the late 1980s through the early 1990s) from a norm-referenced to a criterion-referenced system based on standards represented a significant and fundamental shift in

instructional philosophy. Its implications for teaching and learning cannot be over-emphasized. Hyerle (1996) discussed this fundamental change regarding theories of cognition. He called it a "cognitive revolution" (p. 13), and claimed that we began a slow institutional transformation away from rote behaviorism, closed definitions of intelligence, and the static structure of knowledge. The guiding term for this cognitive revolution in instructional philosophy is Constructivism. The current reform agenda calls for all students to learn, and so requires instructional theory that supports this goal. It is a commonly held belief that the quality of teacher instruction is central to reform goals (Marzano, Pickering & McTighe, 1993; McTighe & Ferrara, 1995; Shepard, 1995; Stiggins, 1988, 1992, 1995, 1996; Wiggins, 1990, 1993). The nature of contemporary, reform-like instruction is aligned with Post-Modern and Authentic philosophies of instruction.

Post-Modernism. Post-Modern philosophy adapts Constructivist learning theories to instructional practice. From this perspective, all knowledge is constructed in the minds of individuals, and teaching embodies a set of values identified as "difference, particularity, and irregularity" (Elkind, 1997). Post-Modern classrooms are student-centered, provide social interactions, inspire individual investigations, value creative expression, and differentiate instruction based on individual learning styles.

Authentic Instruction. As the new standards-based movement was launched across the nation, Newmann and Wehlage (1993) provided five standards of Authentic Instruction, including:

(1) higher-order thinking
(2) depth of knowledge
(3) connectedness to the world
(4) substantive conversation
(5) social support for student achievement.

All of these require a fundamentally different approach to instruction in the classroom. For example, teaching methods related to authentic instruction include using manipulatives and real-life learning opportunities relevant to students' prior experiences. Authentic education's goal is to develop thinking skills for lifelong self-directed learning.

Learning Theory

Cognitive researchers suggest that meaningful learning is reflective, constructive, and self-regulated (Bransford & Vye, 1989; Davis & Maher, 1990; Marzano, Brandt, Hughes, Jones, Presseisen, Rankin & Suhor, 1988; Wittrock, 1991). Studies in cognitive psychology have suggested that students learn better from hands-on, holistic learning experiences (Dietel, Herman & Knuth, 1991). Rote exercises such as structured drills are not effective if the goal is to move students toward higher-level, analytic ways of thinking. Researchers suggest that "to know" something does not mean that a student simply receives the knowledge; it means the student is able to interpret it and relate it to other knowledge. Hands-on, performance-based testing flourished in the early 1990s (Peterson & Knapp, 1993). Although assessment modalities changed, instructional practice was not necessarily altered (Baker, Gratama & Bachtler, 2003).

Constructivism. Constructivist theory proposes that humans construct their own understandings through a cycle of experience and reflection. This cycle is dependent upon the application of cognitive science as new experiences are accommodated by reflection upon their relationships to existing structures. Learners need meaning, conceptual understanding, and real-world connections, not just facts. Constructivist philosophy impacts curriculum, instruction, and assessment. Curriculum must fit each student's individual knowledge base and developmental needs (Sutherland, 1992), and build or accommodate new learning with a hands-on, problem-solving approach. Instruction must incorporate higher-order thought questions to help students make connections, and social dialogue to enable metacognition and analysis of learning. Assessment must be student-centered and offer an array of authentic options. Standardized tests that result in a norm-referenced grading system are not consistent with Constructivist teaching and learning.

Neuropsychology. In their book, *Making Connections: Teaching the Human Brain*, Caine and Caine (1991) examine in detail the way the brain learns. To learn, they suggest, the brain must be involved with "active processing" (p. 147). Caine and Caine describe active processing as "the consolidation and internalization of information, by the learner, in a way that is both personally meaningful and conceptually coherent. It is the path to understanding rather than to simple memory" (p. 147). Active processing assumes a person asks reflective questions about a learning experience: "What did I do?" "Why did I do it?" or "What did I learn?" Neuropsychological research has confirmed that multiple, complex, and concrete experiences are essential for meaningful teaching and learning.

Cognitive theories of learning promoted a shift in educational philosophy and promoted the standards-based reforms that dictate all aspects of contemporary education in America. Lorrie Shepard (1989) summarized this shift in cognitive theory:

> The notion that learning comes about by the accretion of little bits is outmoded learning theory. Current models of learning based on cognitive psychology contend that learners gain understanding when they construct their own knowledge and develop their own cognitive maps of the interactions among facts and concepts [...]. Real learning cannot be spoon-fed one skill at a time. (p. 5)

Put simply, Shepard argues that if we want students to be able to solve open-ended problems and work cooperatively in groups, we should allow students to do so as part of their routine. According to Michaels (1988), "The clear message of reform is that we need to examine our basic philosophical beliefs about teaching, learning, the nature of human beings, and the kinds of environments that maximize growth for teachers and students alike" (p. 3).

Although Newmann and Wehlage (1993) developed their five standards of Authentic Instruction, they also pointed out that research at the time was not definitive about whether or not Authentic Instruction improves student learning more than do traditional forms of Instruction. They recommended that additional research be conducted to determine whether Authentic Instruction produces notable performance effects. Three studies in Washington State did just that, and have

indeed found the links to academic achievement hypothesized by Newmann and Wehlage a decade earlier (Fouts, Stuen, Anderson, & Parnell, 2000; Abbott & Fouts, 2003; Brown & Fouts, 2003). Several studies (Fouts et al., 2002; Abbott & Fouts, 2003; Brown & Fouts, 2003) have revealed strong correlations between student achievement and the presence of Powerful Teaching and Learning in schools. These studies involved more than 10,000 classroom observations over an eight-year period. Powerful Teaching and Learning pulls together all of the theory and practice that aligns with an education system that seeks to educate all students.

Design of the STAR Protocol

During the development of the STAR classroom observation instrument, numerous revisions occurred. These revisions were based on the discussions and critiques of professionals to improve the instruments' language, organization, and format while remaining faithful to the overall tool. For instance, the latest version has added an area to note grade level, status of classroom (ELL, regular, advanced), course level, and number of students.

The STAR Protocol consists of three levels: Essential Components, Indicators, and Strategies. Specifically, five Essential Components build the structural framework Protocol. These include: Skills, Knowledge, Thinking, Application, and Relationships. The five Essential Components are described below followed by sections on Indicators and Strategies.

Essential Components

Skills. This component focuses on skills as the central domain and asks the underlying question; 'Did students actively read, write, or communicate?' The teachers' task is to provide ample opportunity for students to develop and demonstrate skills needed for students to engage in learning. The ultimate goal is for students to develop skills at a conceptual level through such tasks as organizing, creating, analyzing, evaluating, and predicting. Students use various methods and tools to help them reach this goal including use of graphic organizers and visual representations.

Knowledge. This component is organized around the question; 'Did students demonstrate depth of conceptual understanding?' In classrooms scoring high on the component, the teacher provides well-structured assignments and clarity in learning objectives in the classroom to ensure students understand expectations. The premise behind the component is that by allowing students to engage in learning that builds on conceptual understanding, students go beyond simple recall and discover new meaning. The goal is to connect new learning to prior learning, make connections to related information, and generalize this knowledge to new situations so students develop conceptual understanding.

Thinking. The question; 'Did students demonstrate thinking through reflection and/or metacognition?' comprises this component. The teacher is called to teach in a way that provides ample opportunity for the student to develop critical thinking skills. This may mean using well-tailored questions to engage students in classroom interactions and encourage students to develop effective thinking processes. Questions are structured to elicit opinions and to encourage students to support opinions using evidence. This component entails a process of reflection, analyzing, and evaluating one's own thought processes (i.e., self-evaluation, self-monitoring). For instance, receiving effective feedback can provide students with the opportunity to reformulate information or make revisions. The critical thinker uses discussions, opinions, explanations and strategies to reach conclusions, critiques, and constructive feedback. The hope is to have students develop the mental capacity to seek answers and question different phenomena and processes.

Application. The guiding question underlying this component is; 'Did students extend their learning into relevant contexts?' Teachers motivate their students to be more engaged through relating lesson content to other subject areas, personal connections, and contexts. This component seeks to imbue the student with an engaging environment where meaning is attached to on-going experiences by applying theory to practice, relating material to everyday applications, or finding relevance in current, personal, and cultural issues. The component includes increasing students' awareness of personal relevance and the applicability of the curriculum outside the classroom through such avenues as real-life examples, discovery learning, imagery, independent research, presentations, correspondence, and internships.

Relationships. The question connected to this component is; 'Did interpersonal interactions reflect a supportive learning environment?' The goal of the teacher is to provide an ideal teaching and learning environment where students are involved and are excited about the material they are learning. The teacher provides an opportunity for the student to learn in a positive, inspirational, safe, and challenging environment without worry of being judged, criticized, or ridiculed. Knowing students personally, expressing a caring attitude, giving eye contact, and smiling at students are essential elements for fostering meaningful relationships and creating a safe communication climate. The format of the classroom also allows students to engage in collaborative work to encourage their participation. Students learn to generate and respond to information to gain insight from the point of view of others. The learning environment is meant to be supportive in adapting to the needs of diverse learners in hope of reaching those students who learn differently. The goal is to create a classroom climate that enhances students' success by motivating them to succeed.

Indicator Development

Each of the five Essential Components is represented by a set of three Indicators. Some of the Indicators came directly from the TAOP, others came from the Teacher Perspectives Questionnaire, from a rubric created by the Northwest Regional Educational Laboratory, and from the work of Newmann and Wehlage. The Indicators provide ways in which each Essential Component is manifested. The first Indicator in each set focuses on constructivist teacher methods; the following two Indicators focus on student cognitive processes and behaviors that demonstrate learning. This structure illustrates that the STAR Framework is a framework for student-centered work and includes the reciprocal process of teaching and learning, distinct in nature from a teacher-centered instructional framework.

Strategies

Each Indicator includes multiple Strategies for designing Powerful Teaching and Learning experiences. The STAR Framework contains approximately 100 examples of Strategies to manifest each Indicator in a lesson, gathered from teacher reports and classroom observation data. This is not an exhaustive list of the instructional strategies that have proven effective; it is a collection of the most common manifestations seen while conducting primary research.

Validity of the STAR Protocol

After developing the protocol, researchers sought to establish the tool's validity. First, researchers worked to establish content validity by asking a group of twelve educators throughout the state of Washington to review the protocol. This group of reviewers included classroom teachers, building principals, district administrators, education service district personnel, and state level administrators. Reviewers were asked to evaluate the appropriateness of the items for measuring Powerful Teaching and Learning and to provide feedback around how well the protocol aligned with Washington State educational reform goals. Feedback from reviewers was very positive regarding both the appropriateness of the items and the alignment of the protocol with reform goals. Secondly, in order to establish concurrent validity of the STAR Protocol, The BERC Group completed simultaneous scoring within 120 classrooms with the TAOP and the STAR. Regardless of instrument (TAOP or STAR) the overall score remained consistent and the reliability between the two instruments was very high (Fouts, Brown & Thieman, 2002).

Scoring Procedures for the STAR Protocol

The STAR Protocol is designed as a research instrument to identify and measure the degree to which critical elements of Powerful Teaching and Learning are employed and/or are present during any given period of observation time in the classroom. Scoring of the Protocol provides a holistic or Overall score and scores for the five Essential Components and 15 Indicators. A three part process is required for the scoring of the STAR Protocol starting with scoring the Indicators, then the Essential Components and finally giving the lesson an Overall score.

The first step for each observer is to provide a score from 0 to 4 for each Indicator. Each observer uses information gathered from observation of the teacher, student, and the intellectual demands present in the learning environment. The Protocol provides space to mark strategies being used in the observations and write examples and notes pertaining to that particular area. An important aspect of observation using the STAR Protocol is for the observer to only consider what is actually observed during the time period and not to consider, score, or record anything the observer is told that may have preceded or may possibly follow the observation period.

Next, the observer scores each of the five Essential Components from 0 to 4. This involves the researcher evaluating the overall presence of the Essential Component (i.e., skills, knowledge, thinking, application, and relationships. Questions relating to each of the five areas are taken into consideration. These include: *Did students actively read, write, and/or communicate? Did students demonstrate depth of conceptual understanding? Did students demonstrate thinking through reflection and/or metacognition? Did students extend their learning into relevant contexts? Did interpersonal interactions reflect a supportive learning environment?* The researcher also considers a rubric when providing the score that includes percentages of students engaged in developing and/or using skills, constructing conceptual knowledge, demonstrating thinking through reflection or metacognition, extending the context of learning, and demonstrating expectations in the learning environment. Although the observer reviews Indicator scores when rating each Essential Component, the component score is not the average of the three Indicators.

Finally, the observer considers the question, 'How well was this lesson aligned with Powerful Teaching and Learning?' and gives the lesson an Overall score. The score is based on a Likert-like scale of, *Not at all, Very little, Somewhat,* and *Very*. As a general rule, since there are 20 possible points for the sum of the components, the Overall score can be calculated in terms of a range of 0-5 indicating *Not at all*, 6-10 indicating *Very Little*, 11-15 indicating *Somewhat*, and 16-20 indicating *Very*. This general rule is only used as a guide to help determine the extent that Powerful Teaching and Learning is present. For instance, consider if the researcher scores four of the Essential Component areas a 4 and one Essential Component a 0, this would indicate an overall score of 16, and the Protocol Overall score would be a *Very.* However, conceptually this may not be accurate, since one of the Essential Component areas was scored a 0. This finding may reduce the score to *Somewhat*.

DESIGN OF THE STUDY

Observer Training

Observers participated in multiple trainings prior to conducting live observations. The goals of the training included: 1) developing a common understanding and language around Powerful Teaching and Learning; and 2) developing inter-rater reliability for using the instrument. Prior to trainings BERC researchers were provided information on the development of the STAR protocol, the research supporting it, and the connection it has with the other observation protocols such as the TAOP.

A week's worth of training is provided for each observer. Three of the training days consist of viewing videotaped lessons with other observers. These lessons are from a range of grade levels and subject matters. Each observer practices watching a 30 minute video while scanning and taking notes on the STAR Protocol. At the end of the 30 minutes each observer scores the STAR Protocol. After each observer scores the protocol, observers get in groups of two to review and discuss their scoring. Observers specifically spend time in conversations talking about what they saw and why that led them to a particular score. Observers are directed to spend the most time discussing scores that are more than one point away from each other (ex. 1 vs. 3). Agreement is defined as ratings that are within one point of each other. Observer pairs come to a consensus with their scores and present them to the entire group of observers. Discussion and clarification continues when there are any indicator or component scores that continue to be discrepant by more than one point between the observer pairs.

After three days of training using videotaped classrooms, observers shadow a more experienced researcher for at least two full days of live classroom observations. This training begins with

observers understanding how to put together a schedule of classroom observations at a school. Then observers spend the rest of the day scoring the STAR Protocol in live observations with the more experienced researcher. The observer and researcher spend time after each observation to review their scoring and to discuss scores that are not within one point of each other. Periodically, data analysis is completed on inter-rater reliability and Indicators that are consistently not in agreement between observers will be the focus of booster training sessions that are provided on an ongoing basis for all observers.

Selection of Schools and Classrooms

The sample of schools and classrooms represented in this study reflect a sample of convenience taken from a variety of program evaluation projects conducted over the last six years by BERC Group researchers. Over this time period, The BERC Group performed program evaluation projects for the Bill and Melinda Gates Foundation, Washington State Office of Superintendent of Public Instruction, Microsoft, the United States Department of Education, and many districts and individual schools. Some projects include observing all classrooms within a school in order to provide a school-level view of Powerful Teaching and Learning, while other projects entail visiting only certain classrooms in a school to provide a project- or initiative-level view of Powerful Teaching and Learning. For the purpose of this study, researchers created two sample databases. The first database (Database A) includes all classroom observations performed by BERC researchers from the Fall of 2004 to the Spring of 2010. The second database (Database B) includes classroom observations over the same time period performed as part of projects that include visiting all classrooms within a school. Database B will be primarily used for analysis when data is aggregated to the school level.

Schools. The majority of schools in the study sample are from Washington State; however a few schools are included from Hawaii, Missouri, and Oregon. In general, 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. Table 1 shows the comparison between Washington State and Database A and Database B on several demographic variables for six school years.[2]

Table 1. Comparison between State Average and sample Databases A & B

Demographic Variables	**School Year**		
	2004-2005		
	State	**Database A**	**Database B**
# of Schools	2205	51	10
% Male	51.6%	51.5%	51.9%
% Female	48.4%	48.5%	48.1%
% Non-White	29.7%	24.5%	23.8%

[2]To complete Table 1 researchers removed schools outside of Washington State in the sample databases. The number of schools in the sample databases outside Washington State is too small to provide a table comparison with Hawaii, Missouri, and Oregon state averages.

% Qualifying for Free or Reduced-Price Meals	35.9%	37.7%	33.9%
% Qualifying for Special Education	12.3%	11.3%	14.2%
% Transitional Bilingual	7.1%	4.8%	1.2%
Average Years of Teaching Experience	13.1	12.8	13.7
% of Teachers with at least Masters Degree	60.7%	58.8%	57.1%
	2005-2006		
	State	**Database A**	**Database B**
# of Schools	2160	107	44
% Male	51.8%	51.4%	51.1%
% Female	48.8%	48.6%	48.9%
% Non-White	30.8%	39.5%	59.4%
% Qualifying for Free or Reduced-Price Meals	36.7%	51.3%	63.3%
% Qualifying for Special Education	12.4%	NA	NA
% Transitional Bilingual	7.4%	NA	NA
Average Years of Teaching Experience	13.5	13.3	13.7
% of Teachers with at least Masters Degree	61.4%	60.4%	55.8%
	2006-2007		
	State	**Database A**	**Database B**
# of Schools	2321	108	51
% Male	51.5%	51.3%	51.4%
% Female	48.5%	48.7%	48.6%
% Non-White	32.5%	39.7%	54.5%
% Qualifying for Free or Reduced-Price Meals	36.8%	47.4%	52.9%
% Qualifying for Special Education	12.7%	12.7%	13.7%
% Transitional Bilingual	7.5%	9.0%	14.6%
Average Years of Teaching Experience	12.7	12.3	11.9
% of Teachers with at least Masters Degree	61.7%	59.7%	56.4%
	2007-2008		
	State	**Database A**	**Database B**
# of Schools	2365	114	94
% Male	51.6%	51.4%	51.3%

% Female	48.4%	48.6%	48.7%
% Non-White	33.8%	54.0%	55.0%
% Qualifying for Free or Reduced-Price Meals	37.9%	55.7%	58.6%
% Qualifying for Special Education	12.6%	12.6%	12.5%
% Transitional Bilingual	7.9%	13.6%	15.1%
Average Years of Teaching Experience	12.6	11.6	11.6
% of Teachers with at least Masters Degree	63.3%	57.0%	56.5%
	2008-2009		
	State	**Database A**	**Database B**
# of Schools	2429	168	133
% Male	51.5%	51.2%	51.0%
% Female	48.5%	48.2%	48.3%
% Non-White	35.2%	55.0%	57.9%
% Qualifying for Free or Reduced-Price Meals	40.4%	54.0%	58.2%
% Qualifying for Special Education	12.7%	13.3%	13.4%
% Transitional Bilingual	8.0%	12.7%	14.6%
Average Years of Teaching Experience	12.5	11.9	11.6
% of Teachers with at least Masters Degree	64.1%	61.1%	59.6%
	2009-2010		
	State	**Database A**	**Database B**
# of Schools	2410	250	233
% Male	51.6%	51.5%	51.6%
% Female	48.4%	48.1%	48.0%
% Non-White	37.1%	58.8%	59.9%
% Qualifying for Free or Reduced-Price Meals	41.3%	61.4%	63.4%
% Qualifying for Special Education	12.8%	12.3%	12.3%
% Transitional Bilingual	8.1%	15.3%	15.9%
Average Years of Teaching Experience	12.4	11.7	11.8
% of Teachers with Masters Degree	65.8%	60.9%	60.4%

Classrooms. The STAR Classroom Observation Protocol was designed to be used in a variety of subject area classrooms and is not subject-matter specific. Observations are mainly from core subject areas (language arts, mathematics, social studies, and science), but also include observations from classes that are "integrated" or cover multiple subjects areas, physical education classes, career and technical education courses, art classes, and foreign language classes. Typically researchers do not observe high-impact Special Education classrooms.

In general, when conducting classroom observations, BERC researchers send a letter (see Appendix XXX) explaining the classroom observation component of the program evaluation and outlining the general sequence of events. This is followed by a phone call to each school principal to schedule an observation date(s). Each school is asked to provide researchers with a master school schedule, a map, a list of absent teachers, and a list of student teachers (classrooms with substitute teachers or student teachers are not observed). In the morning on the day of the visit, researchers arrive 30 minutes before school starts and put together an observation schedule for the day. Researchers visit classrooms randomly in hopes of seeing typical lessons. In elementary schools, an attempt is made to visit a particular grade level at many different times during the day in order to observe a variety of subject matters. The breakdown of classroom subject matters and grade levels for the two databases is presented in Table 2.

Table 2. Classroom Subject Matters and Grade Levels in Sample Databases A & B

	A	B
Subject Matters		
English	36.0%	37.2%
Math	27.2%	25.3%
Science	13.5%	13.4%
Social Studies	11.9%	12.4%
Integrated	5.2%	5.4%
PE/Health	0.5%	0.5%
Career and Technical	1.2%	1.2%
Art	0.7%	0.7%
Foreign Language	0.7%	0.8%
Other	1.0%	1.1%
Missing	2.0%	2.1%
Grade Level		
Elementary	38.3%	39.0%
Middle/Junior	21.9%	20.3%
High	37.5%	38.5%
Other	1.2%	1.3%
Missing	0.9%	1.0%

General Procedures

The STAR is designed as a research instrument to measure the degree to which Powerful Teaching and Learning ideas are present during any given period of observation time in a classroom. The design of the instrument 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 becomes more important than spending longer periods of time in fewer classrooms. For this reason, the observer is not concerned with what preceded the observation period or what may happen after the observer leaves. 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. By keeping observation time to this range, it is possible to view two lessons during a given class period, although this varies for block schedules. The number of researchers visiting a school depends upon the total number of teachers. In most cases, two to five observers visit a school for one day, with only one researcher observing at a time in any given classroom. In order to check for inter-rater reliability, every tenth classroom is observed by two researchers.

Researchers note grade level, status of classroom (ELL, regular, advanced), course level, and number of students for every classroom observed. Each protocol is coded by number matched to teachers in order to maintain confidentiality. Examples of specific activities, curricular materials, collaborative groupings, classroom climate, and other classroom related information are noted to provide evidence for reports written to the school afterward. 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.

Provisions for Inter-rater Agreement and Reliability Estimates

One of the most critical factors in observational research is the accurate and reliable recording of events as they occur, thus obtaining an objective account of the process cannot be understated in order to control bias and rating errors. Using the TAOP researchers began conducting classroom observation work for a short period of time while calibrating and developing the STAR Observation Protocol. Researchers did this in order to make the STAR more user-friendly and to address some of the limitations of the TAOP. During the development process, researchers used both the TAOP and STAR instruments to score classroom observations. During the course, regardless of the instrument being used, the overall score remained consistent and the reliability between the two instruments was 100% on the overall score.

Training multiple observers and conducting ongoing checks of inter-observer agreement and reliability is needed to minimize observer effects during the observation process. To ensure accurate documentation of classroom events, multiple observers were trained periodically in the use and scoring of the protocol. These trainings consisted of viewing, scoring, and discussing videotapes of various grade and content level lessons. These lessons were observed by the entire team of observers for approximately 30 minutes each. After each observation the team met to debrief, review, and discuss individual ratings for each area of the STAR. During these meetings questions were raised regarding the meaning of particular items and clarification and agreement on the items were sought. This activity was an important aspect of refinement of the process for rating the classrooms.

In addition to training sessions, inter-rater reliability checks were made throughout the development and refinement of the instrument as researchers observed classrooms in a multitude of schools. For approximately every tenth observation visited, two or three researchers observed the same lesson together, scoring the lesson independently. If three or more researchers observed in a school the composition of the calibrated teams rotated in order to ensure the entire team took part in reliability checks. After the visit to the classroom was completed the calibration team would meet to discuss the scores for all areas of the protocol. Inter-rater agreement was reached when observers were within one point of each other.

RESULTS

Inter-rater Agreement and Reliability Estimates

Researchers analyzed approximately one year of STAR Classroom Observation Protocol data to determine the degree of inter-rater agreement and reliability on the Overall score between two raters observing the same classroom and scoring the protocol independently.[3] In total, there were 142 instances where two observers scored the same classroom on the protocol. Inter-rater agreement was calculated by taking the number of instances where observers gave a lesson the same Overall score on the protocol and divided it by the total number of observations with two raters (n = 142).[4] The overall inter-rater agreement estimate for 142 classrooms was 92%. In total, there were 131 instances where observers gave the lesson the same Overall score. In another 10 cases the observer ratings differed by one-point. In only one instance did the observers differ by two-points on the Overall score. The reliability coefficient used for this analysis was Kappa. The Kappa value for this analysis was .90, suggesting very good reliability between observers on the Overall score for the protocol.

Factor Analysis

Researchers employed exploratory factor analysis to investigate the dimensionality of the STAR Classroom Observation Protocol. We used three criteria to determine the number of factors present in the protocol: the presence of eigenvalues over one for a factor, the scree plot, and the interpretability of the factor solutions. Results indicated one dominate factor accounting for a

---

[3] Frick and Semmel (1978) draw a clear distinction between observer agreement and reliability coefficients, "two statistically related but conceptually different indices…"(p. 157). In this study we have chosen to report both indices.

[4] Percentage agreement was calculated using the standard formula (Harrop, Foulkes & Daniels, 1989)

$$\text{Percentage agreement} = \frac{\text{Number of agreements}}{\text{Number of agreements + disagreements}} \times 100$$

While our definition of "agreement" may produce higher percentages of agreement than more strict definitions, we supplement the agreement percentages with a reliability coefficient on the same observations, thus providing a more complete picture of the data.

sizable percent of the variance (see Table 3). The factor analysis indicated an initial eigenvalue of 7.41 for the first factor, which accounted for 49.4% of the variance. The eigenvalue for the second factor was also over one at 1.22, accounting for an additional 8.1% of the variance. Considering the scree plot, eigenvalues, and percentage of variance accounted for, we rotated the first two factors using a Promax rotation procedure and computed an internal consistency estimate of reliability for the 15 Indicators taken together (*Cronbach's Alpha* = .92). Figure 1 displays the scree plot and Table 4 shows the factor loadings for the first and second factors, as well as internal consistency estimates for each Indicator. Factor loadings less than .40 were suppressed for this analysis. Finally, Table 5 displays the factor structure of the STAR.

Table 3. Principal Components: Variance Distribution for Rotated Solution

		Promax Rotation	
Factor	**Eigenvalue**	**% of Variance Accounted For**	**Cumulative %**
1	7.41	49.37	49.37
2	1.22	8.14	57.50

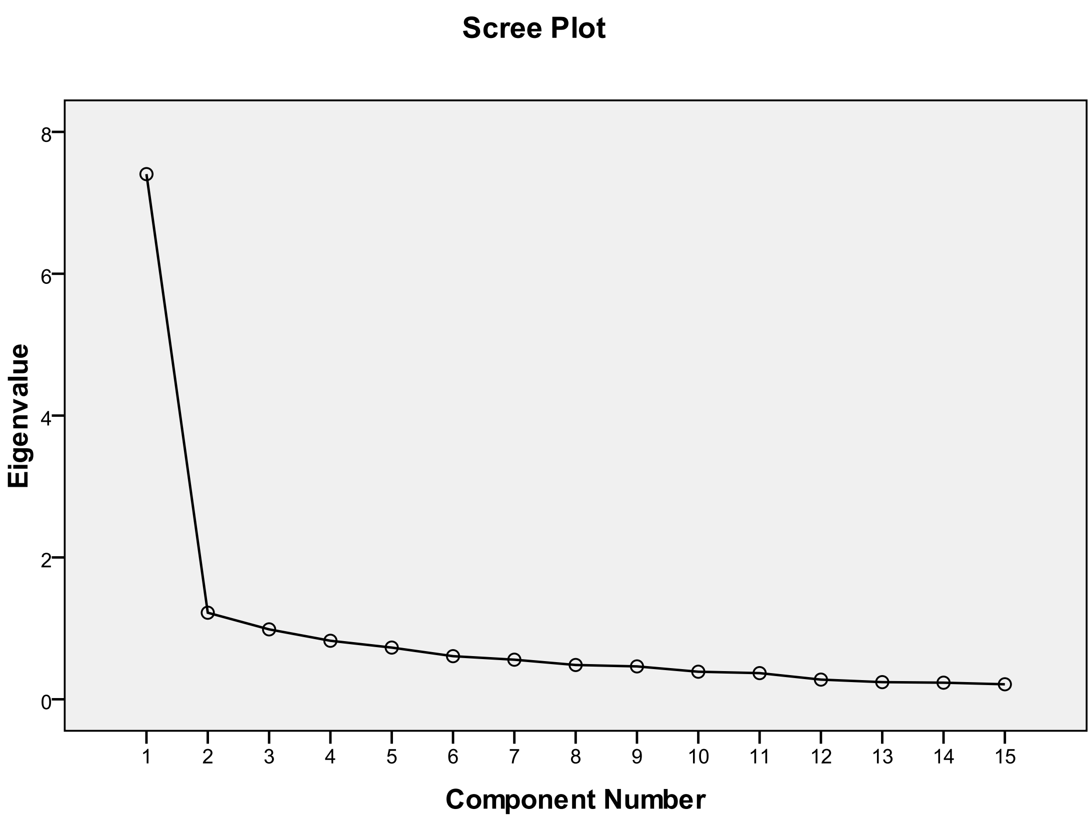

Figure 1. Scree Plot for Factor Analysis of the STAR Protocol

Table 4. STAR Classroom Observation Protocol Factor Loadings (Largest to Smallest) and Internal Consistency Estimates

Factor: *Powerful Teaching and Learning*				**α = .92**
Indicator #	**Indicator Description**	**Factor #1 Loading**	**Factor #2 Loading**	**Alpha if item deleted**
2	Students' skills are used to demonstrate conceptual understanding	.86		.91
5	Students construct knowledge and/or manipulate information and ideas to build on prior learning, to discover new meaning, and/or to develop conceptual understanding, not just recall	.86		.91
6	Students engaged in significant communication, which could include speaking/writing, that builds and/or demonstrates conceptual knowledge and understanding	.86		.91
8	Students develop and/or demonstrate effective thinking processes either verbally or in writing	.84		.91
7	Teacher uses a variety of questioning strategies to encourage students' development of critical thinking, problem solving, and/or communication skills	.79		.91
3	Students demonstrate appropriate methods and/or use appropriate tolls within the subject area to acquire and/or represent information	.78		.91
1	Teacher provides an opportunity for students to develop and/or demonstrate skills through elaborate reading, writing, speaking, modeling, diagramming, displaying, solving, and/or demonstrating	.77		.91
13	Teacher assures the classroom is a positive, inspirational, safe, and challenging academic environment	.74		.91
9	Students demonstrate verbally or in writing that they are intentionally reflecting on their own learning	.72		.91
4	Teacher assures the focus of the lesson is clear to all students	.71		.91
14	Students work collaboratively to share knowledge, complete projects, and/or critique their work	.61		.91
15	Students experience instructional approaches that are adapted to meet the needs of diverse learners (differentiated learning)	.59		.91
12	Students produce a product and/or performance for an audience beyond the classroom		.76	.92

11	Students demonstrate a meaningful personal connection by extending learning activities in the classroom and/or beyond the classroom	.75	.92
10	Teacher relates lesson content to other subject areas, personal experiences, and contexts	.53	.92

Table 5. Factor Structure of the STAR Classroom Observation Protocol

Indicator #	Indicator Description	Factor #1	Factor #2
1	Teacher provides an opportunity for students to develop and/or demonstrate skills through elaborate reading, writing, speaking, modeling, diagramming, displaying, solving, and/or demonstrating	***	
2	Students' skills are used to demonstrate conceptual understanding	***	
3	Students demonstrate appropriate methods and/or use appropriate tolls within the subject area to acquire and/or represent information	***	
4	Teacher assures the focus of the lesson is clear to all students	**	
5	Students construct knowledge and/or manipulate information and ideas to build on prior learning, to discover new meaning, and/or to develop conceptual understanding, not just recall	***	
6	Students engaged in significant communication, which could include speaking/writing, that builds and/or demonstrates conceptual knowledge and understanding	***	
7	Teacher uses a variety of questioning strategies to encourage students' development of critical thinking, problem solving, and/or communication skills	***	
8	Students develop and/or demonstrate effective thinking processes either verbally or in writing	***	
9	Students demonstrate verbally or in writing that they are intentionally reflecting on their own learning	**	
10	Teacher relates lesson content to other subject areas, personal experiences, and contexts		
11	Students demonstrate a meaningful personal connection by extending learning activities in the classroom and/or beyond the classroom		**
12	Students produce a product and/or performance for an audience beyond the classroom		**
13	Teacher assures the classroom is a positive, inspirational, safe, and challenging academic environment	**	
14	Students work collaboratively to share knowledge,	*	

	complete projects, and/or critique their work	
15	Students experience instructional approaches that are adapted to meet the needs of diverse learners (differentiated learning)	*

*Factor Loadings = .55-.65, **Factor Loadings = .66-.76, Factor Loadings = ***.77-.87

In summary, the STAR Protocol consists of one dominate factor accounting for a large percentage of variance. This factor consists of all Indicators on the protocol except 10, 11, and 12 (Application Indicators). A second factor also exists consisting of Indicators 10, 11, and 12. These Indicators relate to the relevancy of a lesson. The internal consistency estimate of reliability *(Cronbach's Alpha*) for the 15 Indicators taken together is .92.

Scoring Characteristics of the STAR Protocol

Attempts to quantify and measure abstract concepts such as Powerful Teaching and Learning are difficult, and as such researchers involved in this study continue to check and discuss their experiences and findings in classrooms. One of the scoring characteristics of the STAR Protocol noted throughout the years of its development is the infrequency of a lesson scoring high on all 15 Indicators of the protocol. Researchers hypothesize that this is due to a couple of reasons. First, there is not always adequate time in a thirty minute observation period for all Indicators of Powerful Teaching and Learning to be utilized (although researchers have noted a few exceptions to this). Second, for any given lesson, not all Indicators are necessarily needed or appropriate. Therefore, researchers noted that a lesson scoring high on Powerful Teaching and Learning might be scored low on a few of the Indicators, but still receive a high Overall score.

A second scoring characteristic of the STAR Protocol is that teacher lecture and discussion approaches to teaching can still provide relatively high scores on the protocol. And conversely, simply because a teacher attempted to use group work or project-based learning, it is not a guarantee of a high score. In fact, the Indicator (#14) related to student collaborative learning has one of the lowest correlations of all the Indicators (11th lowest correlation out of 15) with the Overall score (Pearson $r = .51$). Indicators with the highest correlations to the Overall score are related to students developing conceptual understanding of content (*Knowledge* Essential Component) and students demonstrating effective thinking processes through reflection and metacognition (*Thinking* Essential Component). Indicators within these Essential Components have Pearson r correlations with the Overall score ranging from .62 to .78. Researchers hypothesize that the use of a more traditional, more direct approach to instruction does not preclude a lesson from scoring high on the protocol and in contrast, attempting to use many principles of PTL (ex. collaborative student work etc.), but in ineffective ways can lead to a low Overall score on the protocol. BERC researchers believe that a lesson scoring high on the STAR is less dependent on a specific teaching strategy employed and more dependent on the type of intellectual demands placed on students in a classroom.

Total Sample Scores and Frequencies on the STAR Protocol

The means and standard deviations for the 15 Indicators of Powerful Teaching and Learning used on the STAR Protocol for the entire sample of 14,927 classrooms are shown in Table 6. The means and standard deviations for the 5 Essential Components and Overall score of the STAR are shown in Table 7. The Indicator, Component, and Overall scores were rounded to the whole number and the frequencies of these scores are shown in Figures 2 through 5.

Table 6. Rank Order By Means of the 15 Indicators of the STAR Protocol in 14,927 Classrooms

Indicator #	Indicator Description	Mean	Std. Deviation
13	Teacher assures the classroom is a positive, inspirational, safe, and challenging academic environment	3.2	0.8
1	Teacher provides an opportunity for students to develop and/or demonstrate skills through elaborate reading, writing, speaking, modeling, diagramming, displaying, solving, and/or demonstrating	2.8	1.1
3	Students demonstrate appropriate methods and/or use appropriate tolls within the subject area to acquire and/or represent information	2.6	1.1
4	Teacher assures the focus of the lesson is clear to all students	2.5	1.1
2	Students' skills are used to demonstrate conceptual understanding	2.4	1.1
5	Students construct knowledge and/or manipulate information and ideas to build on prior learning, to discover new meaning, and/or to develop conceptual understanding, not just recall	2.2	1.2
7	Teacher uses a variety of questioning strategies to encourage students' development of critical thinking, problem solving, and/or communication skills	2.2	1.2
15	Students experience instructional approaches that are adapted to meet the needs of diverse learners (differentiated learning)	2.1	1.3
14	Students work collaboratively to share knowledge, complete projects, and/or critique their work	2.1	1.4
8	Students develop and/or demonstrate effective thinking processes either verbally or in writing	2.1	1.2
6	Students engaged in significant communication, which could include speaking/writing, that builds and/or demonstrates conceptual knowledge and understanding	2.1	1.2
10	Teacher relates lesson content to other subject areas, personal experiences, and contexts	1.7	1.4

9	Students demonstrate verbally or in writing that they are intentionally reflecting on their own learning	1.6	1.3
11	Students demonstrate a meaningful personal connection by extending learning activities in the classroom and/or beyond the classroom	0.7	1.2
12	Students produce a product and/or performance for an audience beyond the classroom	0.2	0.8

Table 7. Rank Order By Means of the 5 Essential Components and Overall Score of the STAR Protocol in 14,927 Classrooms

Component Name	Component Description	Mean	Std. Deviation
Relationships	Do interpersonal interactions reflect a supportive learning environment?	3.0	0.8
Skills	Did students actively read/write, and/or communicate?	2.8	1.0
Knowledge	Did students demonstrate depth of conceptual understanding?	2.3	1.0
Thinking	Did students demonstrate thinking through reflection and/or metacognition?	2.2	1.1
Application	Did students extend their learning into relevant contexts?	1.7	1.4
Overall Score	How well was this lesson aligned with Powerful Teaching and Learning?	2.5	0.9

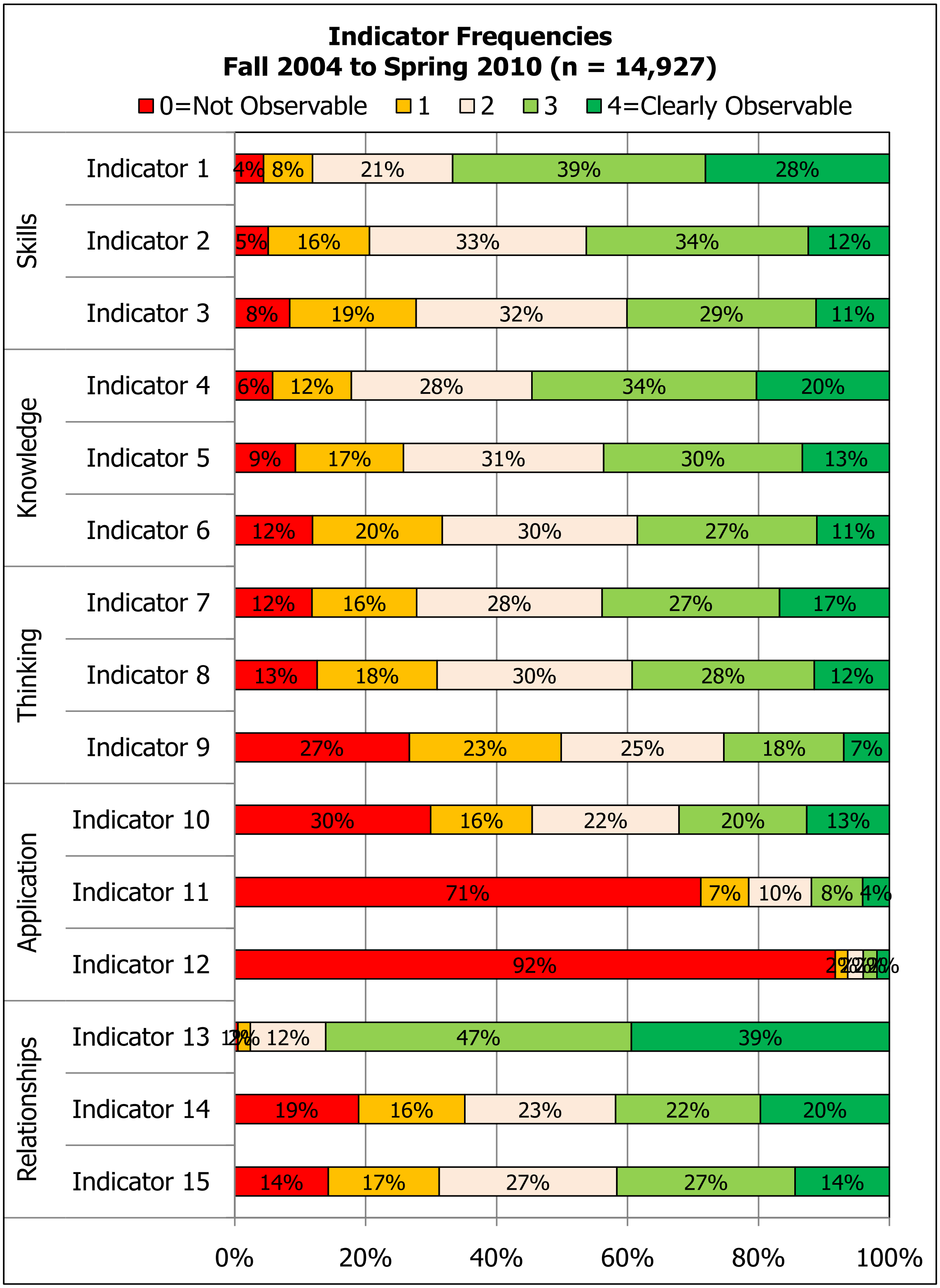

Figure 2. Indicator Frequencies Fall 2004 to Spring 2010

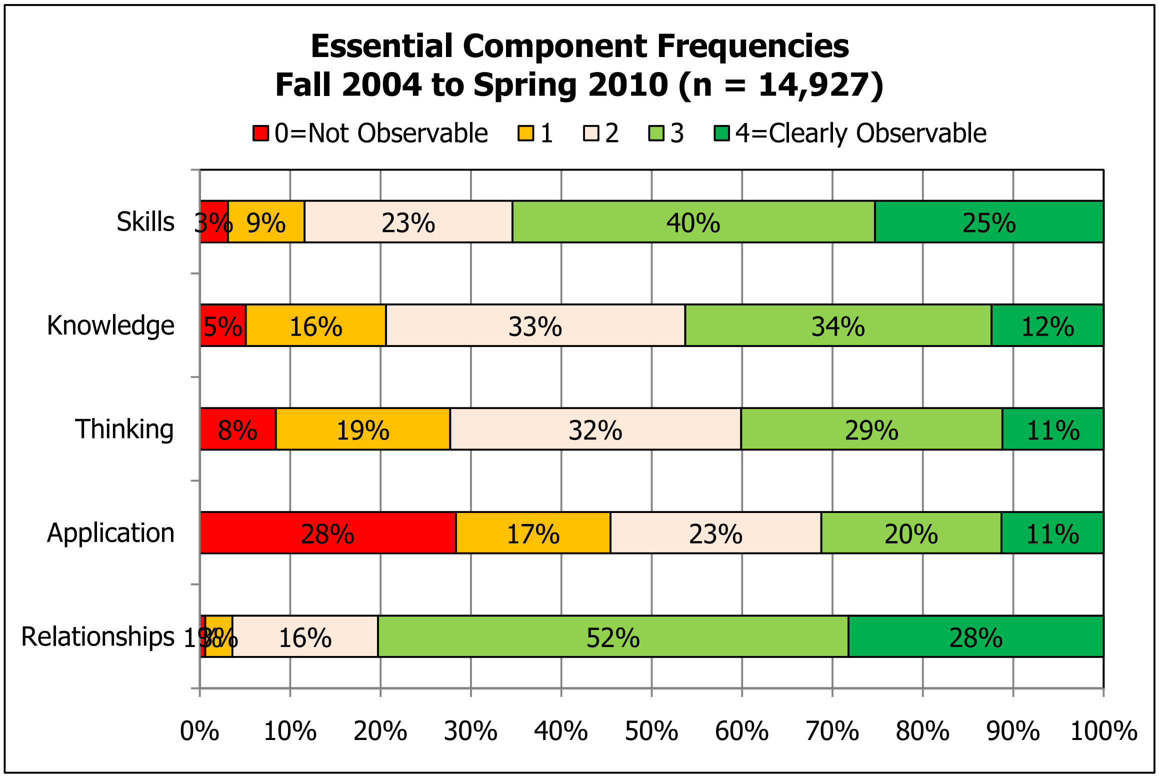

Figure 3. Essential Component Frequencies Fall 2004 to Spring 2010

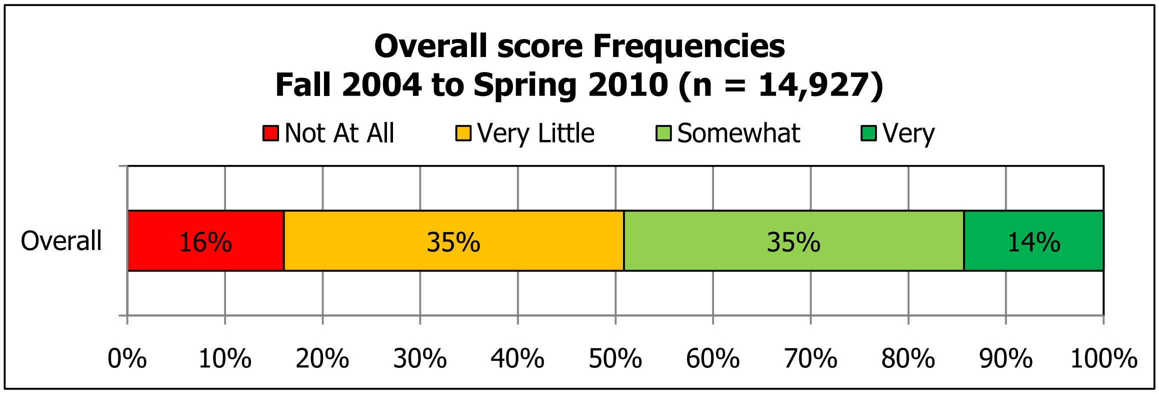

Figure 4. Overall score Frequencies Fall 2004 to Spring 2010

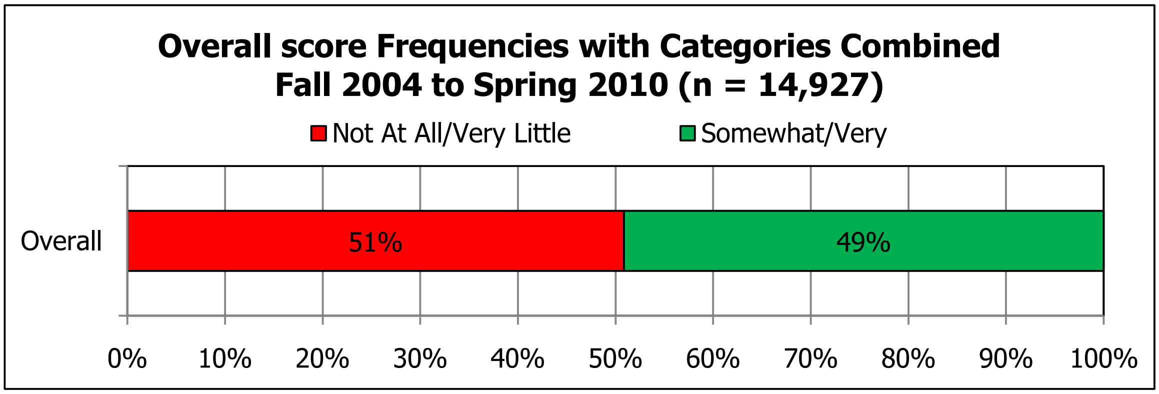

Figure 5. Overall score Frequencies with Categories Combined Fall 2004 to Spring 2010

The data presented in Figure 4 show that only about 14% of the classrooms observed showed very observable evidence of Powerful Teaching and Learning. Another 35% of classrooms observed demonstrated some evidence of PTL. In the remaining 51% of classes, observers saw very little or no evidence of PTL. Of the lessons observed the highest Essential Component rating fell in the area of *Relationships*, followed closely by *Skills* (see Figure 3).The lowest scoring Essential Component is *Application*. These results suggest that most classrooms are supportive learning environments where students are practicing skills, but rarely are students asked to extend their learning into relevant contexts in order to make the learning personally meaningful. Results on the Overall score of the STAR Protocol over the last six years of data collection reveal little change in Powerful Teaching and Learning in classrooms during this time period (see Figure 6). The data used for this chart only includes observations at schools never before observed with the protocol. The percentage of classrooms scoring *Somewhat* or *Very* on the Overall score on the protocol ranged from 38% in 2004-2005 to 50% in 2009-2010. The results of the 1st time observations at schools from Fall 2004 to Spring 2010 combined together compose the STAR Average (see Figures 7 through 9).

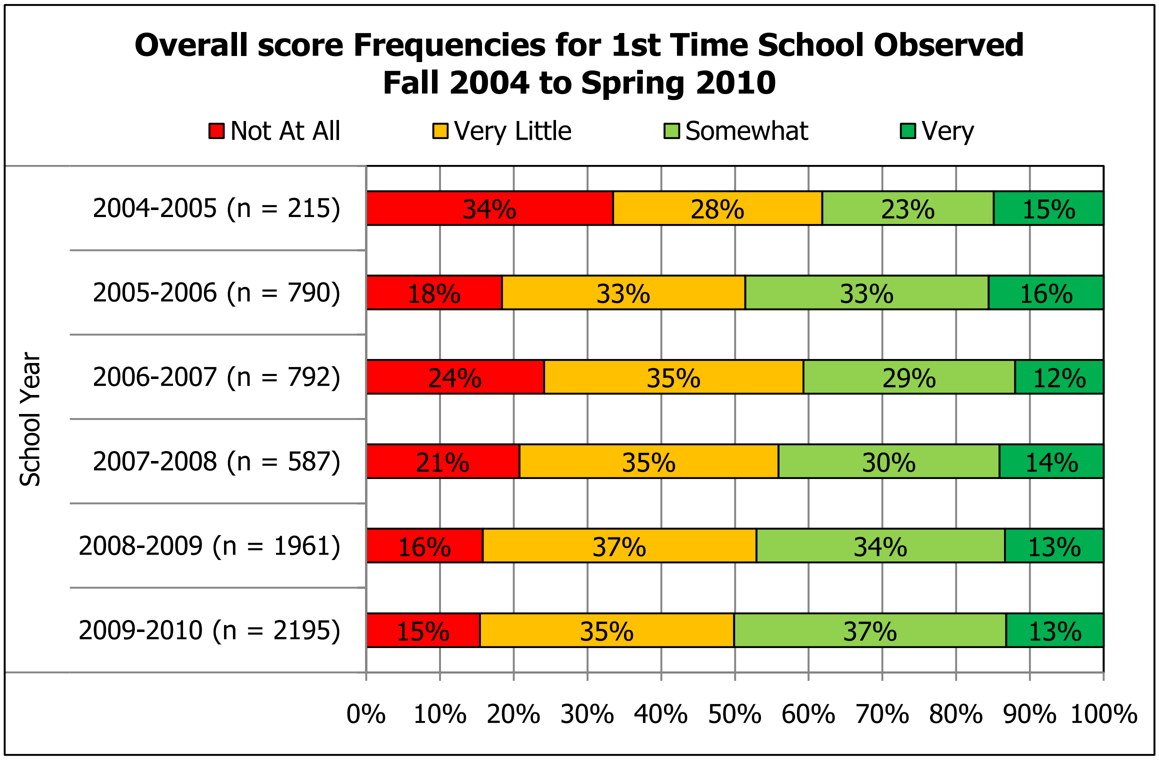

Figure 6. Overall score Frequencies for 1st Time School Observed Fall 2004 to Spring 2010

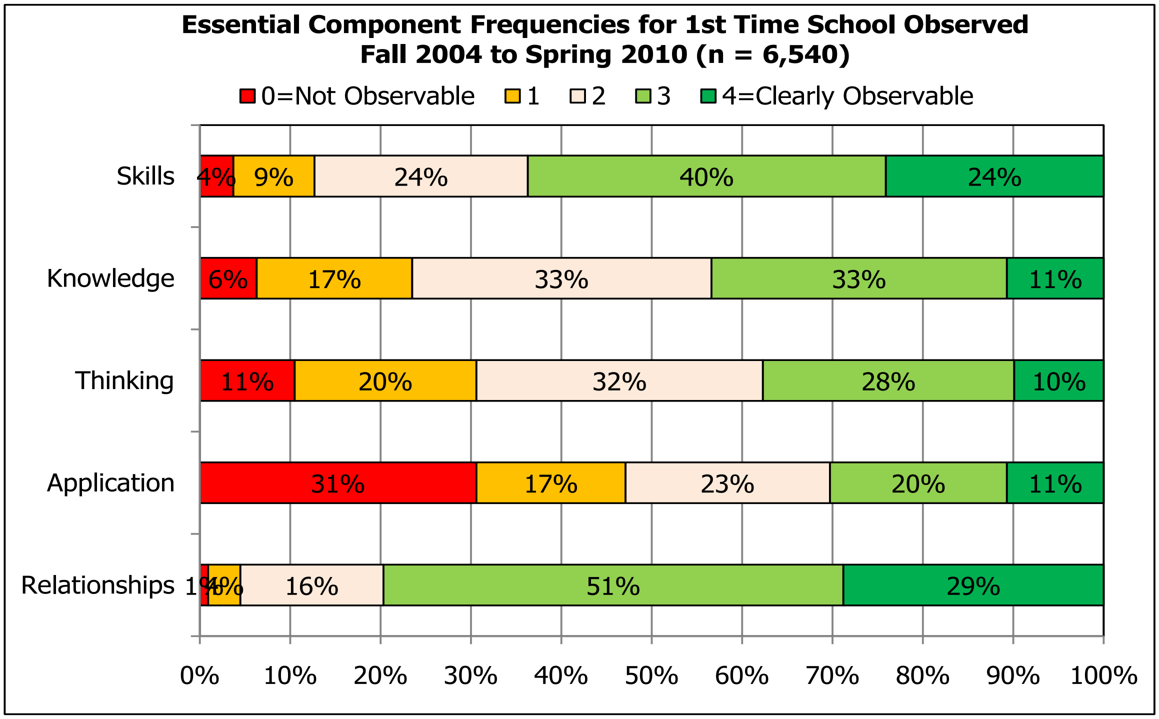

Figure 7. Essential Component Frequencies for 1st Time School Observed Fall 2004 to Spring 2010 (STAR Average)

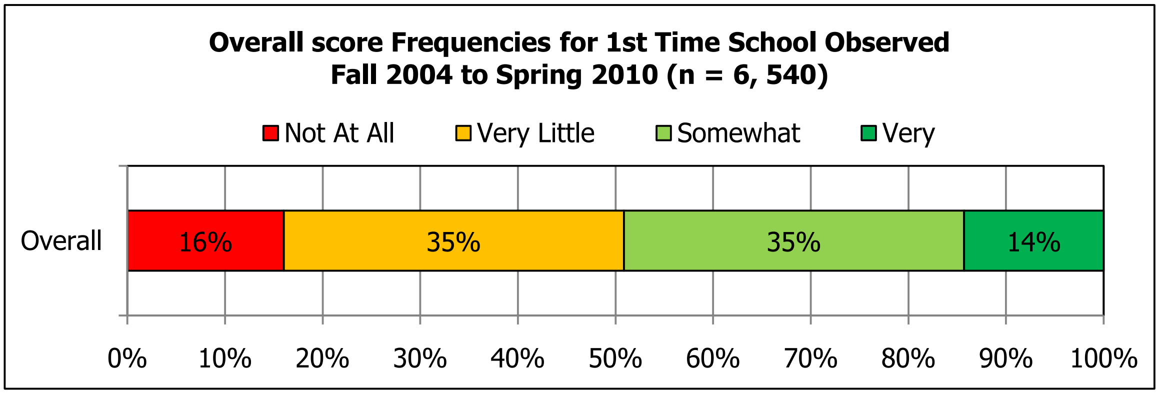

Figure 8. Overall score Frequencies for 1st Time School Observed Fall 2004 to Spring 2010 (STAR Average)

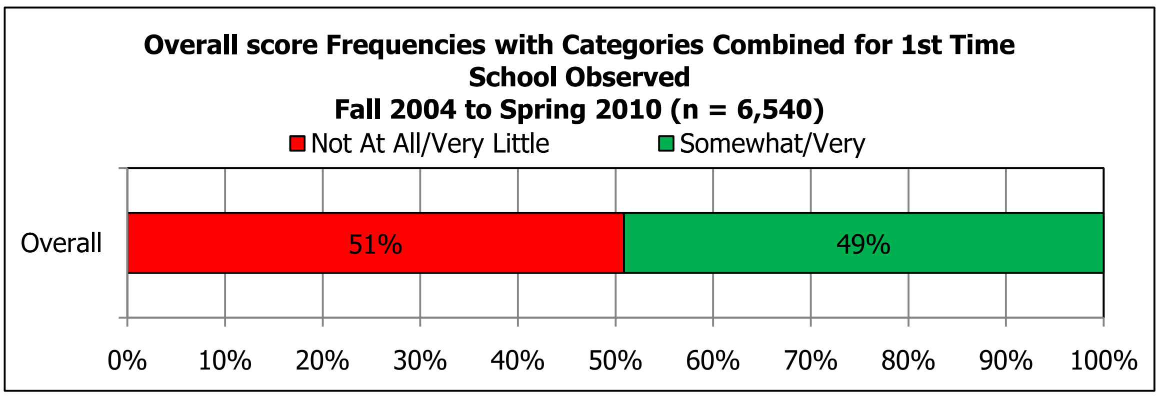

Figure 9. Overall score Frequencies with Categories Combined for 1st Time School Observed Fall 2004 to Spring 2010 (STAR Average)

Pearson *r* intercorrelations among the 5 Essential Components scores and the Overall STAR score are presented in Table 8. All 5 Essential Components correlate significantly with each other and with the Overall score. The strength of these correlations suggests that the instrument reflects internal consistency with the major theoretical construct underlying the protocol.

Table 8. Intercorrelations of the 5 Essential Components and the Overall score

STAR Essential Component	Skills	Knowledge	Thinking	Application	Relationships	Overall
Skills	-					
Knowledge	.70	-				
Thinking	.62	.75	-			
Application	.36	.45	.45	-		
Relationships	.58	.60	.58	.37	-	
Overall	.72	.82	.80	.53	.66	-

Note. All correlations statistically significant at least at the $p < .01$ level (two-tailed).

STAR Scores by School Type and Subject Matter

Means and standard deviations on the STAR were calculated for the entire sample of elementary school classrooms (n = 5467), middle/junior high school classrooms (n = 2838), and high school classrooms (n = 5388). These data are shown in Table 9. The data were also aggregated at the school level, and those means and standard deviations by school type are shown in Table 10. The standard deviations in Tables 9 and 10 are noticeably different. These data indicate that there is considerably more variability among the classrooms than among the schools.

Table 9. STAR Essential Component and Overall score Means and Standard Deviations By Type of School

School Type		Skills	Knowledge	Thinking	Application	Relationships	Overall
Elementary	m	2.95	2.45	2.31	1.65	3.17	2.59
(n = 5467)	sd	.90	.97	1.05	1.29	.71	.89
Mid./Jr. High	m	2.69	2.30	2.10	1.70	2.95	2.44
(n = 2838)	sd	1.05	1.09	1.16	1.36	.84	.94
High School	m	2.56	2.19	1.97	1.74	2.95	2.36
(n = 5388)	sd	1.10	1.08	1.13	1.43	.81	.94

Table 10. STAR Essential Component and Overall score Means and Standard Deviations Aggregated at the School Level

School Type		Skills	Knowledge	Thinking	Application	Relationships	Overall
Elementary	m	2.96	2.46	2.32	1.67	3.18	2.60
(n = 332)	sd	.36	.36	.41	.52	.26	.31
Mid./Jr. High	m	2.69	2.31	2.09	1.68	2.95	2.43
(n = 127)	sd	.38	.36	.38	.40	.29	.26
High School	m	2.58	2.19	1.99	1.76	2.95	2.35
(n = 186)	sd	.39	.35	.41	.47	.32	.28

Means and standard deviations on the STAR were also calculated for different subject matters including English (n = 5349), math (n = 4142), science (n = 2000), social studies (n = 1753), and integrated classes (n = 766). These data are shown in Table 11. Generally, English classrooms had higher means on every Essential Component, except for *Application* where social studies classrooms were the highest. The standard deviations for most components and for the Overall score hover around one except for *Application* where variability is greater and *Relationships* where variability is lower.

Table 11. STAR Essential Component and Overall score Means and Standard Deviations by Subject Matter

Subject Matter		Skills	Knowledge	Thinking	Application	Relationships	Overall
English	m	2.87	2.38	2.22	1.75	3.10	2.53
(n = 5349)	sd	.94	1.01	1.09	1.34	.74	.91
Math	m	2.79	2.34	2.24	1.21	3.02	2.46

(n = 4142)	sd	.94	1.01	1.08	1.27	.81	.92
Science	m	2.67	2.22	2.12	1.90	3.02	2.49
(n = 2000)	sd	1.09	1.08	1.14	1.34	.84	.96
Social Studies	m	2.61	2.11	2.03	2.07	2.93	2.43
(n = 1753)	sd	1.17	1.14	1.21	1.41	.84	.99
Integrated	m	2.75	2.34	1.81	2.02	3.09	2.32
(n = 766)	sd	1.14	1.08	1.11	1.38	.75	.90

Finally, means on the STAR were calculated for the each of the subject areas by school type (see Table 12). In general, these means show mixed results and no clear pattern exists to summarize the data as a whole: however in the case of math and science, for every Essential Component and for the Overall score, elementary school classrooms score higher, followed by middle/junior high school classrooms, and high school classrooms. In some cases, the mean differences are practically significant. For example, the difference between elementary and high school math classrooms in the area of Application is almost half a point.

Table 12. STAR Essential Component and Overall score Means by Subject Matter and School Type

Subject Matter	Grade Level	Skills	Knowledge	Thinking	Application	Relationships	Overall
	ES	2.93	2.47	2.34	1.82	3.14	2.57
English	MS/JH	2.86	2.53	2.37	1.88	3.11	2.58
	HS	2.70	2.41	2.23	1.89	3.03	2.47
	ES	2.95	2.58	2.47	1.49	3.13	2.63
Math	MS/JH	2.91	2.47	2.33	1.41	2.91	2.50
	HS	2.56	2.16	1.99	1.02	2.85	2.20
	ES	3.05	2.79	2.56	2.21	3.28	2.81
Science	MS/JH	2.79	2.55	2.37	2.10	3.03	2.64
	HS	2.53	2.33	2.08	1.74	2.94	2.35
	ES	2.85	2.56	2.39	2.26	3.10	2.60
Social Studies	MS/JH	2.43	2.29	2.02	1.93	2.87	2.34
	HS	2.45	2.30	2.12	2.10	2.92	2.47
	ES	2.83	2.41	2.19	1.74	3.18	2.53
Integrated	MS/JH	3.04	2.56	1.98	2.29	3.00	2.48
	HS	2.85	2.54	2.17	2.48	3.20	2.55

MANOVAs by School Level and Subject Matter

The data from the 13,693[5] classrooms were analyzed using SPSS General Linear Model (Multivariate) with school type as the fixed or independent variable and with the Essential Components and Overall score on the STAR as the dependent variables. The overall results for the

[5] This number is smaller than the total number of observations in the database because only elementary, middle, junior high, and high schools were included in the analysis. K-8 and K-12 schools were not included.

MANOVA were statistically significant, $F = 58.16$, $p < .001$, $eta^2 = .025$. Although the MANOVA yielded statistically significant results, the effect size is very small, therefore these results should be interpreted cautiously. In this case, the sample size is so large that it may be more appropriate to consider the practical significance of the mean differences between the school types. For example, the pattern suggests that elementary school classrooms score highest on the STAR Protocol followed by middle/junior high school classrooms, with high school classrooms scoring the lowest. Generally, this pattern holds true for the Overall score and for all of the Essential Component scores except for *Application*, where high school classrooms score the highest and elementary and middle/junior high school classrooms score similarly.

Follow-up ANOVAs were performed and yielded statistically significant results for all of the Essential Components and for the Overall score. Once again, the results of these analyses should be interpreted cautiously due to the large sample size and the low effect size for all of these ANOVAs. Table 13 reports the F values, p values, and eta^2 for each of the follow-up ANOVAs. Post-hoc tests (LSD) revealed differences between all of the grade levels for *Skills*, *Knowledge*, *Thinking*, and for the Overall score. For *Application*, high school classrooms scored significantly higher than elementary school classrooms, but not different than middle/junior high school classrooms. Additionally, for *Relationships*, elementary school classrooms scored significantly higher than middle/junior and high school classrooms, but middle/junior classrooms did not differ from high school classrooms on this component.

Table 13. Follow-up ANOVA Results

Dependent Variable	*F* Value	*p* Value	*Eta*^{2}
Skills	198.85	< .001	.028
Knowledge	85.93	< .001	.012
Thinking	127.41	< .001	.018
Application	6.14	< .01	.001
Relationships	132.60	< .001	.019
Overall	85.77	< .001	.012

A similar MANOVA was run with subject matter as the dependent variable. The overall results for this MANOVA were also statistically significant, $F = 76.63$, $p < .001$, $eta^2 = .032$. Once again, the effect size is small, so these results should be interpreted cautiously. The sample size is so large that it may be more appropriate to consider the practical significance of the mean differences between the subject matters. For example, the greatest difference in means between subject matters falls in the area of *Application*. On this component, social studies classrooms score the highest and math classrooms score the lowest. The mean difference between these two subject areas on this component is .86, which is almost a full point when thinking about scoring the protocol. While most of the mean differences between subject matters are likely not of practical significance, this one seems to be.

Follow-up ANOVAs were performed and yielded statistically significant results for all of the Essential Components and for the Overall score. These analyses should be interpreted cautiously due to the large sample size and the low effect size for all of these ANOVAs. Table 14 reports the *F* values, *p* values, and eta^2 for each of the follow-up ANOVAs. Post-hoc tests (LSD) revealed differences between most of the subject areas for each of the components and for the Overall score. Some exceptions to this exist, but once again the reader should refer to whether the mean differences between these subjects are of practical significance.

Table 14. Follow-up ANOVA Results

Dependent Variable	*F* Value	*p* Value	Eta^2
Skills	60.13	< .001	.017
Knowledge	18.56	< .001	.005
Thinking	34.75	< .001	.010
Application	196.16	< .001	.053
Relationships	18.71	< .001	.005
Overall	11.26	< .001	.003

STAR Protocol and Achievement

The goal of this analysis is to better understand the relationship between Powerful Teaching and Learning and student achievement. Data for this analysis included school-level Powerful Teaching and Learning scores for 331 schools in Washington State. Each school's PTL score is compiled by aggregating each classroom's Overall score on the protocol. PTL school-level scores are from five school years from 2004-2005 to 2008-2009.[6] For schools observed more than once during a school year, the mean school level PTL score was used.

Researchers collected school level student achievement and demographic data from the Office of Superintendent of Public Instruction state report card website. PTL scores were matched to each school's demographic data and achievement scores. The achievement scores indicate the percent of students who passed the writing, reading, mathematics, and science sections of the state test administered during the same year as the classroom observations. Since the PTL score is representative of the school as a whole, researchers wanted to ensure that the student achievement variable was also as representative of the whole school as possible. So, if schools had more than one grade level taking the state test in a particular subject area, the average percent of students passing was used. For example, in the 2008-2009 school year, 3rd, 4th and 5th grade students took the Reading WASL, so researchers took the average percent of students passing for all three grade levels.

Since previous research by the Washington School Research Center (Abbott & Fouts, 2003) and many others document a strong relationship between low-income and student achievement these

[6] The state test changed in the 2009-2010 school year, so data from that year were not included in this analysis.

analyses control for low-income.[7] Descriptive statistics for each variable are presented in Table 15. Partial correlations among the variables are shown in Table 16. Many statistically significant correlations exist between the variables. Positive and statistically significant correlations existed between PTL and math and PTL and science. There was a trend toward statistical significance with reading. An unexpected negative correlation existed between PTL and writing.

Table 15. Descriptive Statistics for PTL and Student Achievement

Variables	M	SD
PTL	2.45	.30
Reading	65.99	13.86
Math	41.97	15.95
Writing	61.95	17.67
Science	29.94	16.21

Table 16. Correlations between Variables controlling for Low Income

Variables	PTL	Reading	Math	Writing	Science
PTL	-				
Reading	.11^t	-			
Math	.34**	.47**	-		
Writing	-.13*	.50**	-.09	-	
Science	.19**	.31**	.60**	.09	-

$^t p < .06$, $*p < .05$, $**p < .01$

Follow-up regressions were performed using PTL to predict school level student achievement while controlling for low income. In each case, low income was entered as the first variable in the regression equation followed by PTL. Table 17 shows the results of the regression analyses for average school level achievement in each subject area. In each case, the results for the omnibus test (R^2 and *F*-value) are given, along with the contribution of PTL to achievement beyond the effects of low-income. When low-income is accounted for, PTL contributes between 1% and 7% of the variance in achievement. In general, these results indicate positive but small effects of PTL on school-level achievement beyond the contribution of low-income. Specifically, PTL had a strong unique contribution in predicting math achievement beyond the effects of low-income. PTL had a small contribution in predicting writing achievement, but its effects were in the opposite direction of what we would have predicted, implying that schools that have a higher level of PTL, also have lower writing achievement scores.

Table 17. Contribution of PTL to Achievement Accounting for Low-Income

Achievement	Overall R^2 (Adj)	*F*-Value	Unique Contribution of PTL
Reading	.51	168.65***	.01*
Math	.44	127.61***	.07***

[7] The measure of low-income used here is the percentage of students qualifying for free and/or reduced-price meals at a school.

Writing	.32	75.20***	.01*
Science	.48	143.60***	.02**

*$p < .05$, **$p < .01$, ***$p < .001$

Note. Unique contribution of PTL is indicated by the squared semi-partial correlation

SUMMARY AND RECOMMENDATIONS

The purpose of the present study was to describe the development of the STAR Classroom Observation Protocol, examine the psychometric properties of the protocol, examine the extent to which classroom practices are aligned with how students learn best, provide results from six years of classroom observation data, and examine the ability of PTL to predict student achievement. The STAR Protocol is an instrument designed to assess the degree to which Powerful Teaching and Learning is observed in classrooms. The protocol is divided into five Essential Components which include: Skills, Knowledge, Thinking, Application, and Relationships. Three indicators are organized around each Essential Component, comprising a total of fifteen indicators.

A sample of 14,927 classrooms was analyzed to determine observable evidence of Powerful Teaching and Learning. The sample of schools and classrooms represented in this study reflect a sample of convenience taken from a variety of program evaluation projects conducted over the last six years by BERC Group researchers. The majority of schools in the study sample are from Washington State; however a few schools are included from Hawaii, Missouri, and Oregon. In general, 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. Three research questions guided this study: (1) To what extent are classroom practices aligned with how students learn? (2) To what extent have classroom practices changed fundamentally in the last six years? *(3)* Can Powerful Teaching and Learning predict student achievement?

The first research question examined the extent to which classroom practices are aligned with how students learn. Examination of Essential Component frequencies between Fall 2004 and Spring 2010 revealed *Skills* and *Relationships*, respectively, as the Essential Components with the highest ratings. The lowest scoring component is *Application,* followed by *Thinking*. Based on these ratings, the information suggests that classrooms observed are positive, safe, and inviting, providing students with a supportive environment to learn and demonstrate skills. However, students are not being given the opportunity to extend their learning into more relevant contexts very often, nor are students asked to significantly expand on thinking processes to demonstrate and reflect on their learning. Examination of data between Fall 2004 and Spring 2010 of Overall score frequencies showed that only 14% of classrooms observed showed clearly observable evidence of Powerful Teaching and Learning (scoring a *Very*). In an additional 35% of classrooms some evidence of PTL was observed (scoring a *Somewhat*). No evidence or very little evidence was observed in the remainder of the classrooms (51%). This means that only about half the classrooms observed were aligned with how students learn, thus a large percentage of students are experiencing instruction in which they are passively involved in the learning process.

This finding leads into the second research question which examined the extent to which classroom practices have changed in the last six years. The data were based on schools observed for the first-time and are not indicative of what one would expect if PTL was adopted in a school for a period of six years. Overall score frequencies over a six-year span revealed little change in Powerful Teaching

and Learning in classrooms. The percentage of classrooms scoring a combination of *Somewhat* and *Very* on the Overall score ranged from 38% in 2004 – 2005 to 50% in 2009 – 2010, showing an increase of 12%. Although a 12% increase appears to be a moderate amount, the findings are over a period of six years. Therefore, in light of the reform movement to improve instruction, there appears to be some evidence toward this realization, but the process is slow. In general, too few classroom lessons require students to think critically and to relate their knowledge to their own experiences.

The third research question focused on the ability of PTL to predict student achievement. Correlation between PTL and achievement variables (reading, math, writing, and science) shows positive and statistically significant relationships between PTL and math, and PTL and science. The PTL and math correlation was moderate in magnitude. Less significant associations were found for PTL and science, and PTL and reading, which both show correlations small in magnitude. A negative correlation was present for PTL and writing. Regression analysis was conducted as a follow-up to the correlation results to determine whether PTL is a predictor of schools' reading, math, science, and/or writing achievement. The analyses show a positive contribution of PTL to student achievement beyond the effects of income. 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. Only small contributions of PTL were found for reading achievement (1%) and science achievement (2%). A small contribution of PTL was also revealed in predicting writing achievement (1%), however in the opposite direction.

Overall, the most notable findings in this analysis is that Powerful Teaching and Learning predicts additional variance in math achievement beyond the effects of low income. Given the nature of the data (a gross measure of the type of instruction occurring in a variety of classrooms), it does not naturally lend itself to predict school level achievement in a certain subject area by measuring the schools overall PTL score. This is particularly significant considering a restriction of range exists in that there is more variation among classrooms than among schools. In addition, the data shows that Powerful Teaching and Learning is not being used to the extent that reform efforts demand. Approximately half the classrooms observed are not engaged in providing intellectually engaging lessons for students. Furthermore, although a moderate change was found in the extent of classroom practices over a six year period, this change has not occurred at an adequate pace for reform efforts. By drawing upon reform literature, the STAR protocol instrument is established as a model for not only providing a quantitative way for reformed teaching to be measured, but also for showing the extent to which reform has been achieved in classrooms across time.

Align K-12 with College and Work Expectations

Over the past few years there has been increased pressure on the K-12 system and colleges to be more accountable for student learning. Many high school students are not prepared academically for postsecondary studies and enter college without habits or academic experiences of engaging in challenging intellectual tasks (Kuh, 2007). Increasingly, there is a gap between high school and college preparation and work success. Tony Wagner (2008) refers to this as the global achievement gap and suggests that the skills and knowledge taught in high school are not what is needed in today's economy.

Researchers are beginning to recognize the importance of aligning challenging course-taking patterns in K-16 systems as well as the need to develop assessments that align high school exit exams with college entrance exams (Achievers Inc., 2006; Adelman, 1999; Adelman, 2006; Brown & Conley, 2007; Martinez & Klopott, 2005; McDonough, 2004; Venezia & Kirst, 2005). Adelman (1999) stated that college completion is most likely when high school students take more intense and high-quality coursework, especially minority and low-income students. However, more rigorous classes also require instructional methods in higher-order thinking. Brown and Conley (2007) suggest alignment of the high school assessment to university readiness may be limited by less cognitive complex processes. Thus, although these course-taking and alignment aspects are crucial to student success, students are often less prepared in cognitive complex ways.

In rethinking education in light of the reform movement it makes sense that college preparation needs to include high quality instructional practices in high quality courses to provide opportunities for students to engage in higher-order thinking processes. Improving and enhancing instructional practices may create more optimal conditions for more in-depth and sustained learning for all students by providing the means for students to reach the expectations necessary for college and career readiness. Without the inclusion of well-designed instruction, the alignment of K-12 with the college system and the workplace will not be realized fully.

Emphasize Second Order Change

Understanding the differences between types of educational change is required in the post-1993 educational system. In beginning to develop a common language for the change process it is important to consider distinctions between first and second order change. In the educational reform survey work of Fouts, Stuen, Anderson, & Parnell (2000), Baker (1998), Mork (1998), and Van Slyke (1998), schools were asked to identify the focus of their improvement and achievement efforts. Many schools identified First Order Changes such as using specific strategies in their reform efforts (ex. moving to a 4 period day), whereas others identified Second Order Changes that involved a shift in the fundamental beliefs and philosophies. Increases in achievement were identified to a greater degree when schools focused on Second Order Change that involved three constructs: Fundamental Change, Instructional Enhancement, and Collaborative Support (Baker, 1998). According to Fouts (2003):

> There is evidence that one of the reasons schools remain unchanged is that the reforms or changes have been superficial in nature and/or arbitrary in their adoption. Teachers and schools often went through the motions of adopting the new practices, but the changes were neither deep nor long-lasting. In other words, the outward manifestations of the changes were present, but the ideas or philosophy behind the changes were either not understood, misunderstood, or rejected. Consequently, any substantive change in the classroom experience or school culture failed to take root. (p. 12)

Before a First Order Change can be effective, the philosophy and ideas driving practices and actions must be replaced. Second order change alters the underlying philosophical beliefs that in turn drive practice. It is transformational and changes the organization at the core. Oftentimes it involves multilevel and multidimensional changes. Together, first and second order changes help provide a qualitatively different experience for students and raises achievement.

Second-Order Change starts with Fundamental Change and consists of incorporating and embracing a standards-based, criterion-referenced system rather than adhering to the philosophy of the older educational system (Goodman, 1995). It requires schools to fundamentally challenge their belief systems and allow this to drive actions and routines in their work. It consists of understanding the importance of standards and understanding why all students can, should, and will learn. A shift in the philosophy must be the driving force for the change.

A second area of importance is a focus on Instructional Enhancement and requires a commitment to align instruction with practices that help students meet standards. It makes sense that since the classroom is the format where learning takes place the natural focus should be on this area in order for all students to learn. Developing a new and accepted set of ideas or philosophy regarding teaching and learning that leads to a different educational experience for the student involves a conceptual focus. In other words, becoming a more student-centered environment requires attitudinal changes, which are defined as Second-Order Changes. The problem is that since it involves attitudes toward how teachers teach it is a sensitive area. While techniques are fairly easy to change in practice, the attitude toward the change is difficult to alter. While some teachers embrace the change willingly, others may accept it only as something they must do to follow the newest implemented policy, and still others will resist at all costs. The key is to bring on board an enthusiastic group of teachers who can assist in the change process and help change the attitudes of reluctant teachers.

The third area involves Collaborative Support; school personnel working together to ensure all students meet standards. If teachers are to change their instructional practices they must know how to do it. Teachers must know how to provide students with optimal learning opportunities and must depend on widespread support, trust, and collaboration to accomplish it. Ensuring teachers deliver quality instruction requires them to develop a unique set of skills and knowledge. Providing support to teachers is the key to helping teachers move along this pathway. Allowing teachers to observe their peers and learn from each other is an effective approach to promote this type of change in classroom practices. Teachers who are aware of their practices and are able to reflect on their own instruction in a real-life setting with real-life examples grow substantially in the quality of instruction they deliver. Furthermore, it fosters a collaborative culture which sustains improvement by helping teachers develop skills and knowledge continuously.

Current reform efforts involve incorporating into schools all three constructs: Fundamental Change, Instructional Enhancement, and Collaborative Support. Current reform practices are based on a different philosophy and must take place in the classroom. Focusing on instruction and building the capacity and motivation in teachers to improve their instruction through collaborative experiences and by reflecting on their own practices allows them to be effective and efficient in the classroom. Although curriculum and assessment are vital components, without an effective system for delivering the curriculum any changes will have limited impact on outcomes. Reform efforts which fail to address the critical component of instruction and implement this and do not include Second-Order Change are unlikely to deliver the improvements in learning schools are striving to achieve.

Focus on Concepts Rather than Strategies

One characteristic of the observation instrument used in this study is that a lesson scoring high on Powerful Teaching and Learning is less dependent on specific teaching strategies and more dependent on certain types of intellectual demands placed on the student. This finding was similar to other researchers in that it is not the particular teaching strategy employed in the classroom that makes a difference but the quality of the intellectual work the students carry out (Bryk, Nagaoka, & Newmann, 2000; Fouts, 2003; Newmann, Bryk, & Nagaoka, 2001; Newmann, Lopez, & Bryk, 1998; Newman & Wehlage, 1993). No single teaching strategy ensures that students face "high quality intellectual demands" (Fouts, 2003). For example, providing students with 'hands-on' projects or 'collaborative learning groups' does not necessary ensure that a lesson is intellectually demanding for students. In contrast, many lecture-style lessons are observed that are demanding and thought provoking for students. Newmann, Bryk, and Nagaoka, (2001) state:

> Our key point is that it is the intellectual demands embedded in classroom tasks, not the mere occurrence of a particular teaching strategy or technique, that influence the degree of student engagement and learning. Having said this, we do also need to recognize that some teaching practices are more likely to promote complex intellectual work than others. (p. 31)

Along with the current study, Fouts (2003) found that there is no evidence that strong relationships between popular classroom practices alone and student achievement exist. Specific classroom practices and instructional techniques must be accompanied by a new philosophy of classroom instruction. Recent reform efforts are aimed at isolating instructional strategies rather than viewing these as an instructional framework. Many schools have begun to post learning targets on the board to increase the student understanding of lesson objectives. Administrators then engage in 3-7 minute walkthroughs to make sure teachers are posting the targets on the board. While these strategies are worthwhile, they are first-order changes and will produce a limited amount of success. They are inadequate to prepare students for a high standards environment (Fouts, 2003). Focusing on the intellectual demands placed on the students during instruction requires a core modification that alters fundamental beliefs and produces a system based on a different set of guiding beliefs. Second-order changes that encompass a new philosophy and approach to teaching and learning that is needed to reform education.

A Key Approach to Change: An Instructional Framework

Challenging and supporting students of differing abilities and aspirations require a reformed instructional approach to learning. Fouts (2003) suggests, "The most direct impact on student learning is the nature and quality of the classroom instruction and classroom environment to which students are exposed" (p. 16). There are many educational philosophies and perspectives on teaching and learning, and each describe a model of what effective instruction entails. A clear and agreed upon definition of effective teaching and learning is needed.

With the ambition of developing an educational system where *all* students achieve at high levels, the task of defining high-impact instructional practices becomes critical for success. It has become apparent that the Essential Components of Powerful Teaching and Learning provide a framework for addressing this vital need. The PTL framework entails several important elements. One important aspect of this framework requires students to construct knowledge and skills through interpretation and analysis rather than recall. Another aspect is that students are required to engage in thinking processes and be able to articulate these processes. Still another aspect is that students relate information they learn to their personal lives. One final aspect of having a good instructional framework is that the format must include a supportive, challenging, inspirational, and safe environment where students work together as active learners.

For instructional frameworks to be effective they must be adopted into the school system as a whole and not simply practiced in isolated classrooms. The goal is to reach a level of common understanding and practice for all teachers. Just as teachers need a clear understanding of the curriculum they teach, they need a clear understanding of the instructional practices to teach the curriculum. Changing instruction from a teacher-centered to a student-centered classroom does not depend on whether a teacher is born to teach this way. It is an issue of designing lessons and units of study that align with reform efforts. Learning to think in Powerful Teaching and Learning ways will need to be cultivated and clear convictions to the process must be supported. Educators should become a part of a professional learning community where the focus is on instructional practices. They must continue to learn and to develop their own practices through the experiences and support from their colleagues though such activities as studying research, planning lessons together, looking at student work, and observing lessons.

Changes in teaching require changes in the foundational belief system of teachers. Changes such as these follow a constructivist approach and are produced through collaborative experiences in which educators are provided the opportunity to participate in and construct meaning around how children learn best and around what they can do to help with the process (Shulman, 1987). It challenges educators to rethink the teacher and learner roles, and the pedagogy they use to get them there. Simply changing the strategies used to teach without changing the way teacher's think about teaching and learning is insufficient to changing classroom practices. Little progress over this period of time exemplifies the importance of challenging the assumptions of how students learn best and adopting a new set of norms and beliefs about teaching and learning in order to move the reform efforts forward at a quicker pace.

APPENDICES

The STAR Classroom Observation Protocol (STAR)
Letter to Schools
Correlations between STAR Indicators

REFERENCES

Abbott, M. L. & Fouts, J. T. (2003). Constructivist teaching and student achievement: The results of a school-level classroom observation study in Washington. Lynnwood, WA: Washington School Research Center, Seattle Pacific University. Available at: http://www.spu.edu/orgs/research/currentresearch.html

Achieve Inc. (2006). *Closing the expectations gap 2006: An annual 50-state progress report on the alignment of high school policies with the demands of college and work.* Washington, DC: Author. Retrieved September 3, 2010, from http://www.achieve.org/files/50-state-07-Final.pdf (Updated link 7/2007: http://www.achieve.org/files/50-state-07-Final.pdf)

Adelman, C. 1999. *Answers in the Toolbox: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment.* PLLI 1999–8021. Washington, DC: U.S. Department of Education.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, D.C.: U.S. Department of Education. Retrieved Oct 11, 2010 from: HUwww.ed.gov/rschstat/research/pubs/toolboxrevisit/index.htmlUH.

Baker, D. B., (1998). The implementation of alterative assessment procedures and Washington State reform. Seattle Pacific University.

Baker, D. B., Clay J. N., & Gratama C. A. (2005). *The essence of college readiness: Implications for students, parents, schools, and researchers.* Redmond, WA: The BERC Group, Inc.

Baker, D. B., Gratama, C. A., & Bachtler, S. D. (2003). Mathematics Helping Corp: Final report. Olympia, WA: Office of Superintendent of Public Instruction.

Bransford, J., & Vye, N. (1989). A perspective on cognitive research and its implications in instruction. In L. B. Resnick and L. E. Klopfer (Eds.) Toward the thinking curriculum: Current cognitive research. Alexandria, VA: Association Supervision Curriculum and Development.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Brown, R. S., & Conley, D. T. (2007). Comparing state high school assessments to standards for success in entry-level university courses. *Educational Assessment, 12 (2)*, 137-160.

Brown, C. J. & Fouts, J. T. (2003). Classroom instruction in Achievers grantee high schools: A baseline report. Seattle, WA: The Bill & Melinda Gates Foundation. Available at: http://www.gatesfoundation.org/education/researchandevaluation

Bryk, A. S., Nagaoka, J. K., & Newmann, F. M. (2000). Chicago classroom demands for authentic intellectual work: Trends from 1997-1999. Chicago: Consortium on Chicago School Research.

Caine, R. N. & Caine, G. (1991). Making connections: Teaching the human brain. Alexandria, VA: Association for Supervision and Curriculum Development.

Davis, R. & Maher, C. (1990). Constructivist view on the teaching of mathematics. Journal for Research in Mathematics Education. Reston, VA: NCTM.

Dietel, R. J., Herman, J. L. & Knuth, R. A. (1991). What does the research say about assessment? Oakbrook, IL: NCREL.

Echevarria, J., Vogt, M. E., and Short, D. J. (2004). *Making content comprehensible for English learners: The SIOP model.* (2nd ed.). Boston, MA: Pearson Education, Inc.

Elkind, D. (1997). Schooling and family in the post-modern world. In Andy Hargreaves (Ed.), 1997 ASCD year book: Rethinking educational change with heart and mind, (pp. 27-42). Alexandria, VA: Association for Supervision and Curriculum Development.

Fouts, J.T. (2003) A decade of reform: A summary of research findings on classroom, school, and district effectiveness in Washington State. Washington School Research Center. Seattle Pacific University, Seattle.

Fouts, J. T., Brown, C. J., & Thieman, G. Y. (2002). Classroom instruction in Gates grantee schools: A baseline report. Seattle, WA: The Bill & Melinda Gates Foundation. Available at: http://www.gatesfoundation.org/education/researchandevaluation

Fouts, J.T., Stuen, C., Anderson, M.A., & Parnell, T. (2000). The reality of reform: Factors limiting the reform of Washington's elementary schools. Washington School Research Center. Seattle Pacific University, Seattle.

Goodman, J. (1995). Change without difference: School restructuring in historical perspective. Harvard Educational Review, 2, 1-5.

Hyerle, D. (1996). Visual tools for constructing knowledge. Alexandria, VA: Association for Supervision and Curriculum Development.

Kuh, G. D. (2007). What student engagement data tell us about college readiness. *Peer Review, 9* (1), 4-8.

McDonough, P. M. (2004). *School to college transition: Challenges and prospects.* Washington, DC:

American Council on Education.

Marzano, R.T., Brandt, R.S., Hughes, C.S., Jones, B.F., Presseisen, B.Z., Rankin, S.C., & Suhor, C. (1988). Dimensions of thinking: A framework for curriculum and instruction. Alexandria, VA: Association for Supervision and Curriculum Development.

Martinez, M. & Klopott, S. (2005). *The link between high school reform and college access and success for low income and minority youth.* Washington, DC: American Youth Policy Forum and Pathways to College Network.

Marzano, R. J., Pickering D., & McTighe J. (1993). Assessing student outcomes: Performance assessment using the dimensions of learning model. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzono, R. J., Pickering, D., & Pollock, J. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

McTighe, J., & Ferrara, S. (1995, December). Assessing learning in the classroom. Journal of Quality Learning.

Michaels, K. (1988). Caution: Second-wave reform taking place. Educational Leadership, 45(5), 3.

Mork, C.R. (1998). Site-based management and teacher perceptions of restructuring outcomes in Washington State. Published doctoral dissertation, Seattle Pacific University, Seattle.

National Research Council (1999a). How people learn: Brain, mind, experience, and school. Committee on Developments in the Science of Learning. J.D. Bransford, A.L. Brown, and R.R. Cocking (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council (1999b). How people learn: Bridging research and practice. Committee on Developments in the Science of Learning. M.S. Donnovan, J.D. Bransford, and W. Pellegrino (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2002). *Learning and understanding: Improving advanced study of mathematics and science in US high schools.* Washington, DC: National Academies Press.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). The quality of intellectual work in Chicago schools: A baseline report. Chicago: Consortium on Chicago School Research.

Newmann, F.M., & Wehlage, G.G. (1993). Standards of authentic instruction. Educational Leadership, 50 (7), 8-12.

Peterson, P.L., & Knapp, N.F. (1993). Inventing and reinventing ideas: Constructivist teaching and learning in mathematics. In (Gordan Cawelti Ed.) Challenges and Achievements of American Education: The 1993 ASCD Yearbook. Alexandria, VA: Association for Supervision and Curriculum Development.

Piburn, M., Sawada, D., Falconer, K., Turley, J. Benford, R., Bloom, I. (2000). Reformed Teaching Observation Protocol (RTOP).

Shepard, L.A. (1989). Why we need better assessments. Educational Leadership, 46 (7), 4-9.

Shepard, L. (1995). Using assessment to improve learning. Educational Leadership, 52 (5), 38-43.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57, (1),* 1-22.

Simpson, M. L. (2001). *The Art and Science of Professional Teaching: A Developmental Model for Demonstrating Positive Impact on Student Learning.* Education Resource Network.

Simpson, M. L. (2003). *Creating 21st century teachers for Washington State: Linking professional growth to improved student learning.* Olympia, WA: Office of the Superintendent of Public instruction.

Stiggins, R.J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69 (5), 363-68.

Stiggins, R.J. (1992). High quality classroom assessment: What does it really mean? Educational Measurement: Issues and Practice. Reproduced in the Northwest Regional Educational Laboratory training publication: Understanding the meaning and importance of quality classroom assessment (1991).

Stiggins, R.J. (1995). Professional development: The key to a total quality assessment environment. NASSP Bulletin, 79 (573), 11-19.

Stiggins, R.J. (1996). Opening doors to excellence in assessment: A guide for using quality assessment to promote effective instruction and student success. Assessment Training Institutes, Inc. Portland, OR (paper).

Sutherland, P. (1992). Cognitive development today: Piaget and his critics. London: Paul Chapman Publishing Ltd.

Tittle, C. K. & Pape, S. (1996, April). *A framework and classification of procedures for use in evaluation of mathematics and science teaching.* Paper presented at the annual meetings of the American Educational Research Association, New York. Retrieved from ERIC database on September 2, 2010.

Van Slyke, R. (1998). Effects of school reform legislation: The relationship of student achievement gains and degree of restructuring on select Western Washington schools. Unpublished doctoral dissertation, Seattle Pacific University.

Venezia, A. & Kirst, M. W. (2005). Inequitable opportunities: How current education systems and policies undermine the chances for student persistence and success in college. *Educational Policy, 19*(2), 283-307.

Wiggins, G.P. (1990). The case for authentic assessment. American Institutes for Research, Washington, D.C. ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, D.C. p. 4.

Wiggins, G.P. (1993). Assessment to improve performance, not just monitor it: Assessment reform in the social studies. Social Science Record, 30 (2), 5-12.

Wittrock, M. C. (1991). Testing and recent research in cognition. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs: NJ: Prentice-Hall.

The BERC Group, Inc.
22232 17th Avenue Se, Suite 303
Bothell, WA 98021
Phone: 425-486-3100

Web: www.bercgroup.com