2002

Classroom Instruction in Gates Grantee Schools A Baseline Report

Jeffrey T. Fouts

Carol Brown

Gayle Yvonne Thieman
Portland State University, thiemag@pdx.edu

Classroom Instruction in Gates Grantee Schools

A Baseline Report

September 2002

Prepared for the
Bill & Melinda Gates Foundation

Jeffrey T. Fouts
Carol Brown
Gayle Y. Thieman

Fouts & Associates, L.L.C.

EXECUTIVE SUMMARY

Classroom Instruction in Gates Grantee Schools: A Baseline Report

This study is part of the on-going program evaluation of the Bill & Melinda Gates Foundation's Model Schools Initiative and Model Districts Initiative in the state of Washington. In developing the Teaching Attributes Observation Protocol (TAOP) a conceptual framework was identified based on extensive literature on constructivist teaching. From this framework and the foundation's written materials we identified important components and indicators of constructivist teaching and implications for the classroom. We then produced an observation protocol with 7 lesson components and a number of indicators under each component. The content validity of the instrument was then checked against the literature and existing observation instruments.

Following an extensive training period, classroom observations were conducted in 669 classrooms from 34 schools over a four month period of time. Provisions were made for continual checks for inter-rater reliability and agreement, and the results suggest that there was a high degree of consistency in the rating process.

The general findings of this study are that strong constructivist teaching was observable in about 17% of the classroom lessons. The other 83% of the lessons observed may have contained some elements of constructivist teaching, but as many as one-half of the lessons observed had very little or no elements of constructivist teaching present. More constructivist teaching appeared to take place in alternative schools and in integrated subject matter classes than in traditional schools or subjects. There appeared to be no differences among the elementary, middle/junior high and high schools as to the degree to which constructivist practices were used. Finally, the results of the classroom observations do suggest that the there is some validity to the Constructivist Teaching Scale of the *Teacher Perspectives Questionnaire.*

The findings in this study of "constructivist" or "authentic" teaching and learning correspond to other research in the field. Specifically, scoring high on the TAOP as a constructivist lesson is less dependent on specific teaching strategies and more dependent on certain types of intellectual demands placed on the student. This reflects the findings of Newmann, et al. and the Consortium on Chicago School Research that it is the quality of the intellectual work that students undertake that makes the difference.

At the conclusion of the four months of observations, the four observers were gathered together and asked to reflect on their experiences. Their informal observations about the schools and classes, while not research findings in the formal sense, are intriguing and are included at the end of the report.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

We are indebted to a large number of people who made this project possible. Of course, we must acknowledge the contributions of all of those and teachers who opened their classrooms for our observation visits. Without their cooperation, no such research project could be conducted. We also want to acknowledge the hard work and contributions of M.E. Donovan, Amy Rojan, and Susan Gilbert, who, under the leadership of Carol Brown, comprised the observation team and conducted the observations in classrooms all over Washington State.

The following people served as expert reviewers for an early version of the protocol and provided the developer, Gayle Thieman, with critical feedback that was incorporated into revisions of the protocol.

> Christine Chaille is Professor of Curriculum and Instruction in the Graduate School of Education at Portland State University. She is the President-elect of the National Association of Early Childhood Teacher Educators and President of the Oregon Association for the Education of Young Children.
>
> Mark Endsley is the K-16 Assessment Coordinator for the Proficiency-Based Admissions Standards System (PASS) for the Oregon University System. He was the science curriculum specialist in the Gresham Barlow, OR school district and taught professional development workshops in science.
>
> Margit McGuire is the director of the Master In Teaching Program at Seattle University, former president of the National Council for the Social Studies and the Washington Association of Deans and Directors (WACTE). Currently, she is chairing the Pedagogy Assessment Committee for WACTE.
>
> Walter Parker is a professor of education in Curriculum and Instruction, College of Education, at the University of Washington. He has authored numerous articles and texts on social studies curriculum and teaching.
>
> Nicole Miller Rigelman teaches math and K-12 mathematics methods at George Fox University. Previously she was the math curriculum specialist in the Gresham Barlow, OR school district and taught professional development courses in mathematics for the Math Learning Center and Teachers' Development Group at Portland State University.
>
> Carrol Tama is a professor of curriculum and instruction in the Graduate School of Education at Portland State University. She specializes in content area reading and writing with a focus on collaborative learning and reflective practice.

CLASSROOM INSTRUCTION IN GATES GRANTEE SCHOOLS: A BASELINE REPORT

INTRODUCTION: THE CREATION OF HIGH ACHIEVEMENT SCHOOLS

In the year 2000 the Bill and Melinda Gates Foundation announced a $350,000,000 funding commitment to education. "The Foundation's three-pronged investment strategy reflects a commitment to growing successful models that help all students achieve at high levels."[1] The approach includes: (1) recognizing and encouraging high achievement schools and school districts; (2) promoting professional development to enhance district, school and classroom leadership; and (3) helping remove financial barriers to higher education through targeted scholarship programs. Major grants have been awarded in each of these three areas since that announcement.

In the state of Washington 11 five-year district grants and over 50 individual schools grants have been awarded as part of this strategy. All schools within the grantee districts and all schools receiving an individual school grant have been charged with "reinvention" of the school reflecting specific school and classroom *attributes*. These attributes include *Common Focus, High Expectations, Personalized, Respect and Responsibility, Time to Collaborate, Performance Based,* and *Technology as a Tool.* In addition, the schools are charged with improving classroom instruction through the implementation of "powerful teaching" characterized by *Active Inquiry, In-Depth Learning,* and *Performance Assessment.* These "essential components" of powerful teaching have been adapted from *How People Learn: Brain, Mind, Experience, and School* (National Research Council, 1999a) and *How People Learn: Bridging Research and Practice* (National Research Council, 1999b). These essential components are shown in Table 1.

This emphasis on classroom instruction is part of a larger theory of change model for district and school reinvention that has been described in the first year evaluation results for the district and school initiatives (see Fouts, Baker, Riley, Abbott, & Robinson, 2001a; 2001b). This theory of change model for a "standards-based technology-enabled environment" explains the grant program's activities and resources in relation to the intermediary outcomes and ultimate program goals. A full description of the evaluation design is provided in Fouts, et al. (2001a; 2001b). The model allows a multi-level evaluation of the grants at the district, school, classroom and student levels. Baseline assessments of the district practices (attributes), school practices (attributes), and student outcomes were conducted during the 2000—2001 school year. Assessments of the district and school attributes were carried out through interviews and focus groups with educators within the districts and at the schools, and through questionnaires designed around the district and school attributes.

The ultimate goals of the grants are to positively affect student outcomes primarily in the areas of student learning, high school completion, and college attendance. While these student

[1] Quotations in this section and the contents of Table 1 are taken from the Bill & Melinda Gates Foundation website, education division. http://www.gatesfoundation.org/learning/ed/default.htm

outcomes are thought to be influenced to some degree by district functioning and the climate of the school, they are also thought to be most directly influenced by the quality of the classroom instruction that students experience on a daily basis. And so, while districts and schools are expected to change their practices as institutions, they are also expected to facilitate changes in classroom instruction to promote "powerful teaching and learning" characterized by *Active Inquiry, In-Depth Learning,* and *Performance Assessment.* As part of the Year 1 evaluation activities, the teachers completed the *Teacher Perspectives Questionnaire*. Numerous items asked the teachers about the nature of the classroom instruction that takes place at the school, and the degree to which *Active Inquiry, In-Depth Learning,* and *Performance Assessment* are used in that instruction. The classroom observations we conducted in these schools for this study are part of this assessment of grantee practices.

TABLE 1. ESSENTIAL COMPONENTS OF POWERFUL TEACHING

Teachers Focused on Improving Teaching and Learning
The foundation's education grant programs are predicated on three essential components of powerful teaching and learning (adapted from *How People Learn: Bridging Research and Practice*, National Research Council, 1999) in a standards-based technology-enabled environment:

- **Active Inquiry:** Students are engaged in active participation, exploration, and research; activities draw out perceptions and develop understanding; students are encouraged to make decisions about their learning; and teachers utilize the diverse experiences of students to build effective learning experiences.
- **In-Depth Learning:** The focus is competence, not coverage. Students struggle with complex problems, explore core concepts to develop deep understanding; and apply knowledge in real world contexts.
- **Performance Assessment:** Clear expectations define what students should know and be able to do; students produce quality work products and present to real audiences; student work shows evidence of understanding, not just recall; assessment tasks allow students to exhibit higher-order thinking; and teachers and students set learning goals and monitor progress.

We have labeled this "powerful teaching and learning" described in the foundation materials and grantee expectations as "constructivist teaching." While the term "constructivist" or "constructivism" is not used in those materials, we believe that the terms capture the ideas and descriptors used for "powerful teaching and learning." We explore these concepts and their use in the educational literature in much more detail when we explain the development of the classroom observation protocol used in this study.

Purpose of the Study

This study is part of the on-going program evaluation of the Bill & Melinda Gates Foundation's Model Schools Initiative and Model Districts Initiative in the state of Washington and has two purposes. First, the results of the classroom observations reported here will serve as part of the baseline assessment of the nature and amount of *Active Inquiry, In-depth Learning,* and *Performance Assessment* currently being practiced in the classrooms of the Model Districts Initiative schools. Second, the results of these observations will be used as part of a validation study of the *Teacher Perspectives Questionnaire*, which is being used in both the Model Districts and Model Schools grant programs.

DEVELOPMENT OF THE TEACHING ATTRIBUTES OBSERVATION PROTOCOL

In developing the Teaching Attributes Observation Protocol (TAOP), we first identified a conceptual framework from which to base our work. Next, we identified important components and indicators of constructivist teaching from the foundation's written materials, the educational literature, and other observation instruments. This process produced a protocol with 7 lesson components and a number of indicators under each component. At this point six experts with backgrounds in teaching and learning in the areas of mathematics, language arts, science and social studies reviewed the instrument. Their input resulted in a modification of some of the specific indicators, and an affirmation of many of the specifics that had been developed thus far. Final revisions to the instrument took place during four days of training with the four researchers who were going to be doing the actual observations for this study. In this section we describe the theoretical basis for the instrument, the development of the 7 components and 27 specific items, and the establishment of face and content validity. The final version of the TAOP is provided in Appendix A.

Conceptual Framework of the Teaching Attributes[2]

The "essential components" of powerful teaching adapted from *How People Learn: Brain, Mind, Experience, and School* (National Research Council, 1999a) and *How People Learn: Bridging Research and Practice* (National Research Council, 1999b) reflect an approach to learning that has been given considerable attention in recent years. There is a considerable amount of basic research that supports these ideas, and the research has direct implications for how children should best be taught. Collectively, the research has been called the new "science of learning" and is truly basic research in nature. The new science of learning is derived from the findings in developmental psychology, cognitive psychology, linguistics, and neuroscience, and coupled with the philosophical ideas of constructivism (Duffy & Cunningham, 1996). Taken together they serve as the basis for many of the current beliefs about what and how children should learn in school. "Our understanding of human learning has . . . evolved (based on a wealth of evidence collected over a wide range of different domains and media) from a process based on the passive assimilation of isolated facts to one in which the learner actively formulates and tests hypotheses about the world, adapting, elaborating and refining internal models that are often highly procedural in nature" (Shaw & President's Committee of Advisors on Science and Technology, 1998).

The evaluators of the national projects for the Bill & Melinda Gates Foundation have identified various names for this approach, including "authentic instruction, teaching for

---

[2] Certain parts of this section contain material adapted from Fouts (2002), *Research on Computers and Education: Past, Present, and Future.*

understanding, student-centered instruction, and constructivist teaching. Underlying these innovations is the notion of students as active learners and the teachers as guides, or coaches in the learning process" (American Institutes of Research/SRI International, 2002, p.13). They summarize the essential components of constructivism this way.

> The theory of constructivism is based on the idea that people learn better by actively constructing knowledge and by reconciling new information with previous knowledge. The theory rests on several assumptions: 1) some of our notions of what constitutes "knowledge" may be culturally constructed, rather than truth or fact; 2) knowledge is distributed among group members and the knowledge of the group is greater than the sum of the knowledge of individuals; and (3) learning is an active, rather than passive, process of knowledge construction (Conley, 1993). Like current definitions of instruction, constructivism has two components: 1) in the method of delivery (i.e., teaching methods) and 2) in its content (i.e., intellectual quality) (pp. 13-14).

The Coalition of Essential Schools and the National Association of Secondary School Principals are also advocating for instructional changes implied by constructivist ideas.

There are many commonalities in the plethora of material on constructivist theory and pedagogy. For example, Caine & Caine (1991; 1997), Chaille & Britain (1997), and Lambert, et al. (1995) highlight similar major assumptions of constructivist pedagogy, including (a) knowledge exists within the learner; (b) the learner constructs meaning from personal values, beliefs, and experiences; (c) learning is a social activity enhanced by shared inquiry; (d) reflection and metacognition are essential aspects of constructing knowledge and meaning; and (e) learners play a central role in assessing their own learning.

While constructivism as a theory relies heavily on recent cognitive research, educational theorists from Dewey, to Piaget, Bruner, and Vygotsky are credited with laying the foundation for its current form (Brooks & Brooks, 1993; Fosnot, 1996; Lambert et al., 1995). Dewey espoused the social construction of knowledge and the centrality of student experiences for learning. Piaget pioneered studies of the stages of children's cognitive development and originated the principles of assimilation and accommodation to explain learning. His view of learning as constructing and reorganizing knowledge based on experiences, rather than as memorizing information presented by teachers or texts, is the key idea that sets constructivism apart from other theories of cognition (Fosnot, 1996). Piaget elaborated the concept of "reflective abstraction," which is the process by which the human brain generates new possibilities. Piaget hypothesized that when cognitive structures are disturbed by discrepant information, subsequent reflection leads to structural change or accommodation that transforms the original cognitive structure and assimilates the new information. Bruner expanded on Piaget's theories, explaining the role of language and prior experience in the development of cognitive growth (Fosnot, 1996). Vygotsky focused on the effect of social interaction, language, and culture on learning.

Despite the continuing influence of behaviorism on American classrooms, constructivism has continued to evolve through the recent contributions of cognitive psychology and brain research. According to cognitive researchers, knowledge cannot be "transmitted" to students. Building on Piaget's theory, Resnick and Klopfer (1989) argue that learning requires students to elaborate and question what they are told, examine new information in relation to previous understanding, and build new cognitive structures before knowledge becomes "generative" (used to interpret new situations, solve problems, think, and reason).

Reflecting Dewey's emphasis on experience, Bransford and Vye (1989) believe students must have the opportunity to use information to solve problems connected to the real world and experience its effect on their own understanding. Otherwise, students acquire information that remains isolated and can not be applied in new situations. Furthermore, Bransford and Vye argue that students need the experience of "coached practice" in which their attempts at problem solving are supported by their teacher or peers in a cooperative group. This idea of supported learning is an application of Vygotsky's "zone of proximal development."

A number of researchers have developed constructivist principles of learning based on brain research (Caine & Caine, 1991, 1997; Fosnot, 1996; Sylwester, 1995). One of Sylwester's (1995) most interesting premises is that "emotion drives attention, which drives learning and memory" (p. 86). After an extensive discussion of this premise, Sylwester suggests the following active learning strategies which take advantage of emotion and attention in enhancing learning: student projects, portfolio assessments, debates, storytelling, cooperative learning, simulations, role-playing, and metacognitive discussions. All involve the students in actively constructing meaning, based on concrete experiences.

Caine and Caine (1991, 1997) reiterate that the brain is a social organism, shaped by interactions with the environment and interpersonal relationships. Since learning involves conscious and unconscious processes, learners need opportunities for reflection and metacognition. Caine and Caine believe that integrated curriculum, thematic instruction, cooperative learning, and student self assessment are brain-based classroom practices that will improve student learning. Fosnot (1996) incorporates these ideas in her conception of the classroom as "a community of learners engaged in activity, discourse, and reflection" (p. ix).

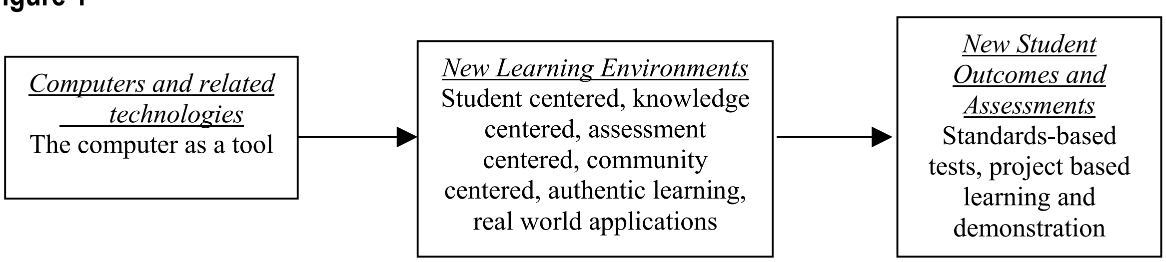
The Role of Technology in Constructivist Classrooms

In the foundation's theory of change model, the goal is to create a "standards-based technology-enabled environment." In this model, technology is seen as a particularly vital component of the reform effort and a means for enhancing constructivist learning. In the past decade the use of the computer and related technologies has expanded from use primarily as an instructional delivery medium to technology as a transformational tool and integral part of the learning environment. In fact, many proponents of the current reform efforts see technology as a vital component of a new educational paradigm in which the curriculum, teaching methods, and student outcomes are reconceptualized (see Means, 1994). This view was adopted by the U.S. Department of Education at least as early as 1993. In U*sing Technology to Support Education Reform"* (United States Department of Education, 1993) it was stated: "technology supports exactly the kinds of changes in content, roles, organizational climate, and affect that are at the heart of the reform movement."[3]

In these settings the computer and related technologies are serving at least four distinct purposes: (1) they are used as previously to teach, drill and practice using increasingly sophisticated digital content; (2) they are used to provide simulations and real world experiences to develop cognitive thinking and to extend learning; (3) they are used to provide access to a wealth of information and enhanced communications through the internet and other related information technologies; and (4) they are used as productivity tools employing application software such as spreadsheets, data bases, and word processors, to manage information, to solve problems and to produce sophisticated products. It is these last three uses that are most important for constructivist teaching and learning.

One of the central components of school reform is the desire for higher academic standards and a stronger focus on higher order thinking, problem solving skills, and learning associated with "real world" applications. To accomplish these ends a new learning environment for schools is necessary. Proponents of school technology assert that it is just that type of environment and those types of learning that are facilitated by the new technology. At the same time there is a predominant belief that the traditional standardized tests are inadequate to measure the types of learning teachers are now being asked to teach. This has resulted in a demand for new assessment procedures for the new learning outcomes. Those new assessments are taking the forms of projects, portfolios, demonstrations, and new standards-based tests. From this perspective technology cannot be viewed or evaluated apart from the other major changes that should take place within the school setting, and is seen as an enabling factor for these other changes. These relationships are shown graphically in Figure 1.

Figure 1



Once again, this changing use of technology reflects the changes in understanding over the last two decades about how the mind works and how children actually learn. The National Research Council's Committee on Developments in the Science of Learning articulated an idea central to this new understanding of human learning: "A fundamental tenet of modern learning theory is that different kinds of learning goals require different approaches to instruction; new goals for education require changes in opportunities to learn" (National Research Council, 1999a, p. xvi). *These new learning opportunities should take place in learning environments*

[3] Many documents found online in non-PDF format do not have page numbers. In this paper page number citations are provided for all hard copy documents in the normal manner. Quotes used without page number citations are from on-line documents with no page numbers.

that are student centered, knowledge centered, assessment centered and community centered, and the new technologies are seen as consistent with the principles of a new science of learning.[4]

Key conclusions of the Committee:

- Because many new technologies are interactive, it is now easier to create environments in which students can learn by doing, receive feedback, and continually refine their understanding and build new knowledge.

- Technologies can help people visualize difficult-to-understand concepts, such as differentiating heat from temperature. Students are able to work with visualization and modeling software similar to the tools used in non-school environments to increase their conceptual understanding and the likelihood of transfer from school to non-school settings.

- New technologies provide access to a vast array of information, including digital libraries, real-world data for analysis, and connections to other people who provide information, feedback, and inspiration, all of which can enhance the learning of teachers and administrators as well as students (National Research Council, 1999a, pp. xviii-xix).

These ideas have been tried by creating technology rich learning environments in basic research settings, not only in the United States, but also in a number of other countries (Vosniadou, DeCorte, Glaser, & Mandl, 1996). In addition, program evaluations over the last decade have shown that when coupled with other factors, the technology can have the effect of creating constructivist environments (see Fouts, 2000).

Implications for the Classroom.

All of these ideas have implications for classroom instruction and set constructivist teaching apart from behavioral or traditional instruction. Fosnot (1996) defined constructivism as a theory about knowledge and learning that describes what knowing is and how one comes to know. She described knowledge as temporary, developmental, internally constructed, and

---

[4] The National Research Council's usage of certain terms in describing these learning environments differs somewhat from the more common usage in education. Learner centered refers "to environments that pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring to the educational setting." It implies "building on the conceptual and cultural knowledge that students bring with them to the classroom"—a basic constructivist perspective. Knowledge centered environments "take seriously the need to help students become knowledgeable by learning in ways that lead to understanding and subsequent transfer." In these environments it is important to identify clearly the domains and knowledge to be learned, including automaticity of skills, but also to help students to develop true understanding. Assessment centered environments provide students with the opportunity "for feedback and revision and that what is assessed must be congruent with one's learning goals." While both formative and summative assessments are important, formative assessments are the assessments vital for enhancing student learning. Community centered environments are where "Students, teachers, and other interested participants share norms that value learning and high standards." The term community includes "the classroom as a community, the school as a community, and the degree to which students, teachers, and administers feel connected to the larger community of homes, businesses, states, the nation, and even the world." A thorough explication of these ideas is provided by the Council in *How People Learn* (1999a), pages 119-142.

socially and culturally mediated. Learning involves struggling with the conflict between the learner's existing model of the world and discrepant new insights. Learning involves constructing a new representation of reality with culturally developed tools and symbols, such as language, and negotiating meaning through cooperative social activity. Therefore, according to Fosnot, certain constructivist principles have direct implications for the classroom.

- Learning is developmental. Teachers need to let learners raise their own questions, generate hypotheses, and test them for validity.
- Disequilibrium facilitates learning. Teachers should offer challenging, open-ended questions.
- Reflective abstraction is the driving force of learning. Teachers should encourage student reflection through journal writing and discussion of connections across experiences.
- Dialogue engenders further thinking. Classrooms should be "communities of discourse" engaged in activity, reflection, and conversation.
- Learning proceeds toward development of cognitive structures. Classroom experiences should focus on central, organizing principles that can be generalized across experiences.

Constructivist-inspired approaches to teaching are prevalent in much of the educational literature. For example, Zemelman, Daniels and Hyde (1998) summarized the principles that underlie the recommendations for practice of national curriculum standards reports, including the National Council of Teachers of English, the National Council of Teachers of Mathematics, the National Council for the Social Studies, and the American Association for the Advancement of Science. They identify a variety of constructivist practices including:

> experiential learning, active student engagement, emphasis on higher order thinking, deep study of fewer topics, student meta-cognition, increased student choice and self evaluation, cooperative, collaborative activity, reading whole books and primary sources, performance-based assessments, scoring rubrics/guides, and student portfolios. (pp. 5-15).

Additionally, constructivist pedagogical practices have become an important component of standards-based reform. This connection has been articulated by Wolf et al., (1991, pp. 47-48) as they define the "epistemology of mind" that underlies standards-based teaching and assessment. All learners construct, rather than merely absorb, knowledge through inference, observation, rule generation, and theory building. Learning is understanding how to apply what one knows, not just the amount of information one can absorb. Knapp (1997) explained that standards-based teaching and learning emphasize the learners' understanding of central processes and ideas, the students' ability to reason and apply ideas to non-routine complex problems, and an in-depth immersion in themes and topics rather than superficial coverage of curriculum. In her study of changes in classroom practice in response to standards-based reform in Washington middle schools Thieman (2000) found the development of the standards movement also contributed to the application of constructivist principles.

The foundation's teaching attributes of *Active Inquiry, In-depth Learning,* and *Performance Assessment,* along with the further descriptors shown in Table 1, are clearly reflective of the constructivist/authentic instruction theories of teaching and learning. The

Teaching Attributes Observation Protocol was developed from this theoretical framework and is the basis for this study.

Item/Component Development

A major consideration during the development process was the desire to create a generic instrument to be used across the academic disciplines, and across the elementary, middle, and high school levels. The initial development of the Teaching Attributes Observation Protocol (TAOP) was based on the three components of powerful teaching and learning (see Table 1) identified by the foundation and adapted from *How People Learn: Bridging Research and Practice* (National Research Council, 1999a). The initial protocol was organized around the three components of active inquiry, in-depth learning, and performance assessments, with many of the descriptive statements used to provide operational definitions of the components. Each of the indicators was further defined with examples of what these indicators would look like in language arts, social studies, science, or mathematics classrooms in intermediate, middle school, and high school classrooms.

Subsequently the observation protocol was revised to focus on the areas of seven of twelve items on the Teacher Perspectives Questionnaire (see Fouts, Baker, Riley, Abbott, & Robinson, 2001a; 2001b) developed for the evaluation of the district and school grants in the state of Washington. The items were selected based on factor analysis results and as representative of constructivist practice as defined in the literature. The questionnaire items were based on the three teaching attributes and related descriptions, and therefore many of the ideas from the initial development activities were simply reorganized around the seven items, which became the major components of the protocol. The seven components defining constructivist teaching are listed below:

1. Student work shows evidence of understanding, not just recall.
2. Students are engaged in activities to develop understanding and create personal meaning through reflection.
3. Students apply knowledge in real world contexts.
4. Students are engaged in active participation, exploration, and research.
5. Teachers utilize the diverse experiences of students to build effective learning experiences.
6. Students are presented with a challenging curriculum designed to develop depth of understanding.
7. Assessment tasks allow students to exhibit higher-order thinking.

The next step was the development of more detailed indicators under each of the seven components. Two specific sources of information and five published observation instruments were influential in this process. First, the standards for authentic instruction and assessment developed by Newmann, Secada, and Wehlage (1995) were considered. These standards focus on construction of knowledge (higher order thinking, organization of information and consideration of alternatives), disciplined inquiry (deep knowledge of disciplinary content and processes, elaborated written communication, and substantive conversation), and value beyond school (problems connected to the world beyond the classroom and an audience beyond the school). Second, the observation protocol developed by Thieman (2000) for Washington State middle schools was particularly significant because of its constructivist theoretical base and its use in the state in which the study was to be conducted.

In addition to the larger review of literature on constructivism and these two particular sources, five formal constructivist classroom observation protocols were also reviewed. A brief description of these protocols is provided below.

- *The Accelerated Schools Project Classroom Observation Notes* (Accelerated School Project, 2001). Designed for use at multiple grade levels/subject areas. Focuses on "Organizing Instruction for Powerful Learning" consisting of five components: authentic, interactive, learner-centered, inclusive, and continuous. Specific items on the ASP protocol are similar to the seven indicators of constructivism used in the TAOP.

- *The PATHWISE R Classroom Observation System* (ETS, 2000). Designed to evaluate the instructional practice of student teachers and first-year teachers and is based on a constructivist view of teaching and learning. Two of the domains ("Organizing Content Knowledge for Student Learning" and "Teaching for Student Learning") are particularly relevant to the TAOP.

- *The Reformed Teaching Observation Protocol* (RTOP) (Arizona Collaborative for Excellence in the Preparation of Teachers, 2000). Designed to measure the degree to which mathematics or science instruction is 'reformed' and embodies the standards developed by the National Council for the Teaching of Mathematics, the National Academy of Science, and the American Association for the Advancement of Science. The 25 indicators on the RTOP were designed to measure the "inquiry orientation" of the instructional practices being observed and were influential in the development of the TAOP.

- *The Classroom Observation Scales* (Secada and Byrd, 1993). Designed to assess mathematics instruction in nine areas: use of mathematical analysis, depth of knowledge and student understanding, mathematical connections, cross-disciplinary connections, value beyond the class, mathematical discourse and communication, locus of mathematical authority, social support, and student engagement in mathematics.

- *The CETP Classroom Observation Protocol* (CETP Core Evaluation, 2001). Developed by researchers at the College of Education and Human Development, University of Minnesota, to rate science and math instruction in K-12 classes. The instrument draws on the RTOP as well as the work of Newmann, et al. (1995). The CETP protocol was found to be compatible with the TAOP.

Based on these sources, an early version of the protocol with seven components and 28 specific indicators was reviewed individually by six experts in teaching and learning, and with expertise in mathematics, language arts, science, or social studies instruction. Their responses were generally favorable, but questions were raised about specific items and the appropriateness of one instrument to assess instruction in multiple disciplines. Based on feedback from these people, several items were reworded, and two items were combined to reduce the number of specific indicators to 27. Wording of the indicators was revised further during the training session with observers described in the next section. The seven components and 27 indicators in the final version of the observation protocol are shown in Table 2.

TABLE 2. TAOP 7 COMPONENTS and 27 INDICATORS OF POWERFUL TEACHING

I. STUDENT WORK SHOWS EVIDENCE OF CONCEPTUAL UNDERSTANDING, NOT JUST RECALL

1. Students use appropriate methods and tools of the subject area to acquire and represent information.
text analysis, creative or expository writing, discussion, oral presentation, reading, interviews, desktop publishing, manipulatives, models, maps, timelines, calculators, primary sources, drawing, graphs, symbols.

2. Students develop conceptual understanding.
organizing information, applying information, considering alternatives, interpreting or evaluating, predicting, comparing, contrasting, analyzing cause & effect, hypothesizing, sequencing, developing a model, simulation, or original creation.

3. Students demonstrate thinking by using vocabulary and fundamental concepts of subject area.
literary genres, cause and effect, chemical properties, number theory, probability & statistics.

4. Students construct knowledge by manipulating information and ideas to solve complex problems, discover new meaning, and/or develop understanding.
analyzing a story, discussing a public issue, using historical evidence or current data to support an opinion, analyzing an environmental problem, using symbolic representation, theory building where appropriate.

5. Students communicate conceptual understanding through elaborated writing, speaking, modeling, diagramming or demonstrating.
poetry, essays, journals, research papers, letters, response logs, lab reports, dialogue, debate, skit, presentation.

II. STUDENTS ARE ENGAGED IN ACTIVITIES TO DEVELOP UNDERSTANDING AND CREATE PERSONAL MEANING THROUGH REFLECTION

6. Students use an appropriate learning strategy to gain meaning.
graphic organizer, mapping, drawing pictures, outlining, creating a model, journaling, discussion, reference to text.

7. Students rethink (revise) work based on data, self-evaluation and/or constructive feedback from peers/teacher.

8. Students consider alternatives and/or multiple ways to investigate and problem solve.

9. Students intentionally reflect on their own learning (metacognition).
text to self, other texts, world connections, examining own bias or opinion, critique science lab procedures, math reasoning.

10. Teacher provides focused feedback and questions to students that probe students' conceptual understanding and lead to sense making.

11. Students and/or students and teacher engage in substantive conversation which builds knowledge and develops critical thinking.
literature circle, readers' theatre, discuss writing process, simulation, town meeting, debate, generate hypotheses, share and compare results discuss conclusions, math reasoning.

III. APPLY KNOWLEDGE IN REAL WORLD CONTEXTS

12. Teacher or Student connects knowledge to relevant personal experiences.

13. Teacher or Student connects knowledge within or across disciplines or to a real world problem.

14. Instruction uses community resources or data.
guest speakers, materials.

15. Students produce a product or performance for an audience beyond the class.
persuasive essay, speech, play, posting student work to a website, letter to the editor, pen pals, brochure, community survey.

16. Students interact with world outside school via field-based experiences or technology.

IV. STUDENTS ARE ENGAGED IN ACTIVE PARTICIPATION, EXPLORATION AND RESEARCH

17. Students work collaboratively to share knowledge, complete projects, and./or critique their work.
writing, response partners, reading groups, research groups, lab groups, math problem solving groups.

18. Students generate their own ideas, questions, or hypotheses.

19. Students plan and/or carry out independent research.
choose research topic, information sources, design lab procedures and search for math patterns.

20. Students independently access/use print media, equipment or technology.
books, newspapers, maps, graphs, charts.

V. TEACHER USES DIVERSE EXPERIENCES OF STUDENTS TO BUILD EFFECTIVE LEARNING

21. Teacher activates and accesses prior knowledge of students.

22. Student needs and strengths are accommodated through differentiated learning.

23. Lesson builds on diverse cultural traditions, student interests and experiences.
writing connected to student experience and knowledge, diverse literature; interview family members, lab activities incorporate personal experience, multiple perspectives on numeric.

VI. STUDENTS ARE PRESENTED WITH A CHALLENGING CURRICULUM DESIGNED TO DEVELOP DEPTH OF UNDERSTANDING

24. Lesson presented emphasizes conceptual understanding, not just recall or superficial understanding.
comprehension, analysis of literature, support thesis with data, (re) discover theory, math problem solving.

25. Central ideas and concepts of the subject are covered in depth.
comprehension, continuity/ change, compare/contrast, cause/effect, number theory, measurement, probability, matter, properties, interdependence

VII. SUMMATIVE ASSESSMENT ALLOWS STUDENTS TO EXHIBIT HIGHER ORDER THINKING AND CONSTRUCT KNOWLEDGE

26. Assessment requires Students to communicate learning through elaborated writing, speaking, modeling, diagramming, or demonstrating.

27. Assessment criteria focus on demonstration of knowledge and conceptual understanding of core concepts.

Content Validity of the Teaching Attributes Observation Protocol

To examine the content validity of the protocol, the 27 indicators were aligned with the main ideas of a small sample of authors from the literature review, and aligned with the elements of several existing constructivist observation protocols. Those alignments are shown in Table 3. In the table an “X” in columns 2-5 indicates that the idea, activity or practice expressed in the indicator corresponds to the examples or definitions of constructivist teaching used by that author. An “X” in columns 6-9 indicates that the idea, activity, or practice expressed in the indicator corresponds to similar elements in other constructivist protocols.

TABLE 3. CONTENT OF THE TAOP IN RELATION TO CONSTRUCTIVIST LITERATURE AND INSTRUMENTS

	Constructivist Literature				Constructivist Observation Protocols			
TAOP Indicator (See Table 2)	**Fosnot**	**Bransford & Vye**	**Caine & Caine**	**Zemelman, et al.**	**Newmann , et al.**	**RTOP Arizona**	**ASP Class. Notes**	**ETS Pathwise**
Indicator #1				X	X	X	X	
Indicator #2				X	X	X		
Indicator #3				X	X	X	X	
Indicator #4	X	X		X	X	X	X	
Indicator #5	X			X	X	X	X	
Indicator #6				X	X	X		
Indicator #7			X	X	X	X	X	
Indicator #8	X			X	X	X		
Indicator #9	X	X	X	X		X		
Indicator #10	X	X		X	X	X	X	X
Indicator #11	X			X	X	X	X	
Indicator #12	X			X	X	X	X	X
Indicator #13	X	X		X	X	X	X	X
Indicator #14				X	X		X	
Indicator #15				X	X			
Indicator #16				X			X	
Indicator #17		X	X	X	X	X	X	
Indicator #18	X			X	X			
Indicator #19	X			X	X		X	
Indicator #20				X			X	
Indicator #21				X	X	X	X	X
Indicator #22		X		X			X	X
Indicator #23				X			X	X
Indicator #24			X	X	X	X		
Indicator #25	X			X	X	X		
Indicator #26				X	X			
Indicator #27					X	X		

Scoring Procedures for the Teaching Attributes Observation Protocol

A complete copy of the TAOP is provided in Appendix A. The TAOP is designed as a research instrument to measure the degree to which constructivist teaching and learning ideas are being employed and/or are present during any given period of observation time in a classroom. The scoring of the TAOP generally consists of three steps. The first step involves general observation of student and teacher activities and the nature of student work or the intellectual demands being placed on students. The protocol provides room and encourages observers to take notes to "describe the lesson, the classroom setting, classroom environment, resources, content or skills taught, teacher and student activities, and student work displayed." Observers are encouraged to, "if possible, look at assignments, project directions, or assessments in which students are involved during your observations." An important procedure for the TAOP is for the observer to *only consider what was actually observed during that period of time* and to not record or score the lesson based on what the observer was told preceded or followed the observation period.

The second step is the numerical scoring of the 27 indicators (Table 2) of constructivist teaching and learning. At the conclusion of the observation period observers rate the lesson on each indicator on a 0 to 4 scale for the degree to which the indicator was descriptive of the lesson. The numerical scores represent "never occurred," "occurred very little," "occurred somewhat," "occurred quite often," and "very descriptive." If the observation period did not contain a summative assessment activity, indicators 26 and 27 are given an N/A (not applicable). If the observation period was primarily a summative assessment activity, *only* indicators 26 and 27 receive a rating.

The third step of the scoring is giving a holistic rating to the observation period. Observers are asked to respond to the following prompt:

Overall Conclusion: How constructivist was this Lesson?
Not at all Very Little Somewhat Very

Therefore, the scoring of the observation period provides both an analytical and a holistic score.

DESIGN OF THE GATES GRANTEE STUDY

Observer Training

The observation team consisted of four individuals, all of whom were former classroom teachers; two with experience in elementary school, one with middle school experience, and one with high school experience. In addition, three of the four had considerable university teaching experience in schools of education. The observers participated in four days of training prior to the beginning of the observations. The goals of the training were three-fold: (1) to develop a common understanding of constructivist practice; (2) to critique and revise the protocol; and (3) to develop inter-rater reliability when using the instrument.

At the beginning of the first day of training, a review of the context and background of the study was provided to the observation team. The observers then developed a list of elements of constructivist practice as they understood it. This list was then compared to the TAOP, which indicated that the team had addressed all seven indicators of the protocol. A presentation and discussion of the protocol items generated considerable conversation about how each item might manifest itself in a language arts, math, science, or social studies classroom. The observation team then watched a videotape of a fifth grade social studies lesson from the PASS Project (NCSS, 2001) and scored ten of the protocol items. Individual ratings by the team members were discussed for each item, and several items on the protocol were revised for clarification. The team watched and scored a second videotaped lesson (fourth grade social studies lesson) with subsequent discussion, clarification, and slight revision of the protocol.

On day two, after viewing, scoring, and discussing a videotape of an elementary math/art lesson, the team traveled to an elementary school in the Seattle area for a live pilot test of the instrument. Four classes were observed by the entire team for approximately 45 minutes each: fourth grade math, fourth grade writing, third grade reading, and fourth grade reading. After each observation the team met to debrief, review and discuss individual ratings by the team members for each item on the TAOP.

Prior to day three, specific items on the protocol were revised to provide clarification of meaning by including specific classroom examples of the observed behavior. For example, item 3 (Students demonstrate thinking with vocabulary and fundamental/core concepts of the subject area) was modified to include ". . . such as literary analysis, cause and effect, chemical properties, number theory, probability and statistics." Item 5 was changed ("Students communicate learning through extended writing, speaking, modeling, diagramming, or demonstrating") to read "Students demonstrate conceptual understanding through extended writing (poetry, essays, journals, research papers, letters, response logs, lab reports), speaking (dialogue, debate, skit, presentation), modeling, diagramming, or demonstrating. Such clarifications resulted in increased agreement by individual team members on the ratings for each observation. In addition, indicators six and seven were also modified. "Challenging curriculum which developed depth of understanding" was changed to "Students are presented with a challenging curriculum designed to develop depth of understanding." This change encouraged team members to focus on the presented curriculum rather than inferring the curriculum from the activities. "Assessment tasks allow students to exhibit higher order thinking and construct knowledge" was changed to "Summative assessment allows students . . ." and the scoring guide was changed to include NA (Not Applicable) rather than 0 (Never Occurred). Finally, an overall rating was created for "How constructivist was this lesson?"

On day three, the team reviewed the revisions to the protocol and watched and scored three videotaped lessons: seventh grade social studies, high school chemistry, and high school English literature. Once again, the team discussion after scoring each lesson revealed questions about constructivist teaching, as well as growing agreement about the meaning of each item on the protocol.

The fourth and final day of training occurred at a middle school in the Seattle area. Four classes were observed by the entire team for approximately 30 minutes each: sixth grade math, seventh grade social studies, fifth grade language arts, and seventh grade science. After each observation the team met to debrief, review and discuss individual ratings by the team members for each item on the TAOP.

By the fourth day the team had viewed and fully scored a total of twelve lessons ranging from third grade to eleventh grade in language arts, math, science, and social studies. During the training sessions the calculation of inter-rater agreement was based on the percentage of items on which the four observers agreed. "Agreement" was defined as ratings that were within one point of each other on the five point scale for indicators 1 through 27. Inter-rater agreement for scoring of the videotaped lessons was 69% and for the classroom lessons was 75%. Generally, inter-rater agreement for lessons that were scored very low as constructivist lessons was higher than for lessons that were scored as somewhat constructivist, for which there was less agreement. The total percentage of agreement for all 12 observations was 73%. The specific information for each observation and scoring during training is presented in Table 4.

TABLE 4. SUMMARY OF OBSERVATION RESULTS DUIRNG TRAINING AND FIELD TESTING

Day	Mode	School level	Subject	# of items on which observers agreed (27 items total)	% agreement (all 4 raters scored item within one point of each other)
2	video	E	Math/Art	21	78%
2	class	E	Math	19	70%
2	class	E	LangArts	23	85%
2	class	E	LangArts	22	81%
2	class	E	LangArts	21	78%
3	video	MS	SocSt	12	44%
3	video	HS	Science	20	74%
3	video	HS	LangArts	21	78%
4	class	MS	Math	17	63%
4	class	MS	SocSt	19	70%
4	class	MS	LangArts	23	85%
4	class	MS	Science	19	70%

Selection of Schools and Classrooms

Schools. In the planning stages of the project, we recognized that one of the limitations of this type of comparative research is the tendency for schools to exhibit more within school variance than between school variance. To provide the largest variability between schools on the variable of interest, constructivist teaching, we used teacher perspectives on the amount of constructivist teaching at the schools for the selection process. The results of the Constructivist Teaching Scale on the *Teacher Perspectives Questionnaire* from the previous school year were rank-ordered and used for school selection by elementary, middle/junior high, and high school status. Approximately equal numbers of the highest and lowest scoring schools on this scale were selected for the study, resulting in 15 elementary schools, eight middle/junior high schools, nine high schools, and two technical schools. If selected, participation in the observation study was a required evaluation activity by all schools in the Model Districts and Model Schools Initiatives. The schools and their characteristics are provided in Appendix B.

Classrooms. The *Teaching Attributes Observation Protocol* was designed to be used in a variety of subject area classrooms and is not subject-matter specific. Our observations were conducted in the required core academic courses or elective academic courses in the language arts, mathematics, science, or social studies areas. The observations also took place in classes that were "integrated" academic classes, particularly at the elementary level.

Initially, each district coordinator received a letter (See Appendix C) explaining the classroom observation component of the evaluation and outlining the general sequence of events. This was followed by a phone call to each school principal to schedule an observation date(s). In some cases the principal chose to coordinate the site visit, while at other schools the responsibility was delegated to a building coordinator or librarian. Whatever the case, each school was asked to provide a master school schedule so that an observation plan could be developed. This allowed observers to select classrooms at random in the hopes of seeing typical lessons. For middle and high schools, scheduling was based on number of sections of core classes (language arts, math, science, and social studies) in an effort to adequately represent the curriculum, while in elementary schools, an attempt was made to schedule observations at times when core subjects were being taught. The breakdown of the schools and subject matter classes observed is provided in Appendix D.

While most principals were extremely helpful in providing the research team with information and help in organizing the observations, there were four schools where the contact persons were hesitant and/or unwilling to allow random access to classrooms. In most cases their hesitation was due to pressure from teachers who were reluctant to be observed. When this situation arose, the coordinator/principal provided a prepared schedule for observers. While this limited the ability of observers to see "random" lessons, the observers did feel like they saw a fairly representative sample of classrooms, and therefore likely did not significantly change the overall score of the building.

General Procedures

The TAOP is designed as a research instrument to measure the degree to which constructivist teaching and learning ideas are being employed and/or are present during any given period of observation time in a classroom. This makes the instrument somewhat different than observation tools used for instructional evaluation and improvement for a given teacher. For this use, observers are generally interested in being able to place the observational results in a larger context of a longer instructional unit to get a more complete picture of a single teacher's instructional approach or expertise. In contrast, the intent of the TAOP is to measure what is going on in a given period of time for generalizing not to a single teacher, which would take more or longer observations, but to the school in the aggregate. Therefore, sampling a larger number of classes becomes more important than spending longer periods of time in fewer classes. For this reason, the observer is not concerned with what preceded the observation period or what may happen after the observer leaves. The observer only records and scores the nature of the classroom activities during the period of time she/he is in the classroom. Although the instrument may be used over longer periods of observation time, it may also be used for shorter periods, and in this study observation periods were 30 minutes. This shorter period of time increased the possible number of observations in a school on which to base a general view of the school's instructional practices.

In most cases, two to four observers visited each school for one or two days, depending on the size of the school. The classroom observations were conducted by one observer in each classroom, with approximately every tenth classroom observed by two or more observers for inter-rater reliability checks. Grade level, subject area, and time of day were noted, as were specific activities, curricular materials, student groupings and the like. At the end of 30 minutes, observers scored all 27 items, and calculated an average score for each of the seven components of the protocol. Scores were assigned based only the events that occurred during the 30 minute time period. The class session was also given a holistic score of 1 to 4.

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. Obtaining an objective account is essential, and as such, the importance of selecting and training observers cannot be overstated. While it is unlikely that observer effects, such as bias and rating errors, can ever be completely eliminated, they can be controlled to a large extent through training, the use of multiple observers, and by conducting ongoing checks of inter-observer agreement and reliability. All of these strategies were used in this study to ensure the collection of reliable data.

Training. To ensure accurate documentation of classroom events, observers were trained over four days in the use and scoring of the protocol as described above. This was the first step toward inter-rater reliability. In addition to the training conducted prior to the beginning of the study, there was a continual process of "debriefing" and discussion throughout the four months of observations. These sessions consisted of on-going conversations about the nature of constructivist teaching at the various grade levels and in the various subject matter areas. These activities were an important part of the refinement and clarification of the process of rating the classrooms.

Multiple Observers. Accurate documentation of events can also be accomplished, in part, by utilizing multiple observers. In this study, four observers were used to collect data. All four were former classroom teachers, two with experience in elementary school, one with middle school experience, and one with high school experience. In addition, three of the four had considerable university teaching experience in schools of education.

On-going Reliability Checks. In addition to the reliability work during the initial four-day training session, reliability checks were made throughout the four months of observations in 32 of the 34 schools. For approximately every tenth classroom visited, two, three, or all four researchers observed the same lesson together and then scored the lesson independently. Because a two-person team often visited a single school, the composition of the teams was rotated to insure reliability checks took place on a regular basis among all four observers. The inter-rater agreement rates and reliability estimates from these joint observations are presented in the Results section below.

RESULTS

Inter-rater Agreement and Reliability Estimates

During the months of October through January, a total of 669 classrooms from the 34 schools were observed. In 73 of these classrooms two or more observers were present and scored the classes independently. Inter-rater agreement and reliability estimates[5] from these observations were then calculated.

The Teaching Attributes Observation Protocol uses a five-point scale on six of the seven components (0= Never Occurred to 4 = Very Descriptive). The scoring for the seventh component (Summative Assessment) has an "NA" (Not Applicable) option rather than a "0" (Never Occurred) option, to allow for situations where a summative assessment was not part of the observed lesson. Calculations of inter-rater agreement were based on the percentage of items on which observers agreed. "Agreement" was defined as ratings that were within one point of each other.[6] Thus, a lesson where observers agreed on 26 of 27 items received an agreement estimate of 96%. In the few instances where agreement fell below an acceptable level of 85%, observers discussed differences until they concurred on the scoring. These instances provided the opportunity to "correct" a particular observation's score, but equally important provided the opportunity for on-going discussions about the scoring procedures throughout the research period. The *overall* inter-rater agreement calculations were based on the original agreement percentage and not on the rescore. The average inter-rater agreement estimate for all 73 classroom observations was 93%.[7] The average inter-rater agreement estimate was 92% for high

[5] 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.

[6] The definition of "agreement" is a subjective one. In their discussion of the question, "Agreement Under What Conditions and How 'Perfect?'" Frick and Semmel (1978) examine the position of Medley and Norton (1971) on the possibility, or even desirability, of perfect inter-rater agreement. According to Frick and Semmel, Medley and Norton argue "that perfect observer agreement during actual data collection may *not* be particularly desirable. Since teachers and pupils in the real world do not always exhibit behaviors that neatly fall into predefined observational system categories, observer disagreement on ambiguities reveals a more representative picture of that real world"(p. 162). Frick and Semmel point out that this position seems problematic from a reliability standpoint, but that "it does have merit from a practical standpoint." They base this on the fact that it seems very unlikely that in any study there will be measured constructs that that fit perfectly into "mutually exclusive categories," but are rather ambiguous. They explain: ". . . if an observer codes an ambiguous behavior into one category and another observer codes the same ambiguous behavior into a different category, the overall results may indicate a more realistic description of the behavior of that teacher . . ." They conclude that "disagreement on ambiguous events in the field" should be expected (p. 180). While our definition of "agreement" may produce higher percentages of agreement than the more strict definition, we supplement the agreement percentages with corresponding reliability coefficients on the same observations, thus providing a more complete picture of the data.

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

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

As they point out, however, this formula has been criticized because the percentage agreement can be inflated by chance, leading to an unwarranted "form of investigator's self-awarded 'seal of approval'" (p. 182). They identify several formulas for calculating the percentage chance agreement, but also point out that there is no agreement on school classrooms, 96% for middle school classrooms, and 95% for elementary school classrooms. The results for all of the paired classroom observation inter-rater agreement calculations are shown in Appendix E.

In addition to the percent agreement, 73 separate reliability coefficients were calculated for the 28 ratings given each of the joint observations (27 indicators and the holistic rating). The median inter-rater reliability was .84, and the mean inter-rater reliability was .82. The estimated score reliability[8] for the rest of the classroom observations is .69.

Scoring Characteristics of the Teaching Attributes Observation Protocol

Attempts to quantify and measure abstract concepts such as *constructivism* are difficult, and as such the observers involved in this study continued to check and discuss their findings throughout the course of the observation period. One of the scoring characteristics that became evident during the observation process was that a strong constructivist lesson as defined by the literature would seldom, if ever, score high on all of the 27 individual criteria, and would also seldom, if ever, score high on all seven of the major components of the protocol, particularly Component Seven—Summative Assessment. This was due to at least two reasons. First, there simply was not adequate time in a thirty minute observation period for all of the components of "powerful teaching" to be utilized. Second, for any given lesson, not all of the components were necessarily needed or appropriate. Therefore, it became evident that a strong constructivist lesson might be scored low on several of the 27 criteria, but still receive a high holistic score.

This scoring characteristic of the protocol became most noticeable when the holistic score (1-4) for the lesson was compared to the averages of components 1 through 6 of the protocol. Lessons given a high holistic rating (seen as strong constructivist lessons) of "4" might

---

how much observer agreement should exceed chance agreement. They conclude that, although a number of summary mathematical techniques have been developed, the methods produce conflicting results and none are any more satisfactory than the standard method. They state, "There is obviously an awareness that overall percentage agreement can mislead, but there appears to be no consensus alternative ways of examining data" (p. 184). . . . "There is apparently no best method for computing observer agreement" (p. 188). In lieu of such a method, they recommend presenting contingency tables showing numbers of agreements and disagreements on occurrences and non-occurrences. However, given the number of joint observations (73) we have chosen to simply report the standard percentage agreement supplemented with the reliability coefficients.

[8] Fan and Chen (2000) point out that inter-rater reliability coefficient computed on a small portion of a sample of observations reflects the reliability of that sample but cannot be generalized directly to the remaining portion of the observations for which there was only one rater. This is because the coefficient is based on the *mean* of two observers' scores, and when just one observer is used, the score reliability will inevitably be lower. "Statistically, such average (or total scores) across two raters tend to be more stable (i.e. reliable) than scores provided by only one rater" (p. 533). It, therefore, becomes necessary to estimate the reliability of the scores of the remaining observations based on the inter-rater reliability coefficient. Either classical reliability theory or generalizability theory can be used in this instance because only one source of error (rater inconsistency) is involved. We have chosen to calculate the estimated score reliability for the one rater observations using classical reliability theory and the generalized Spearman-Brown formula:

$$r_{2X} = \frac{k\, r_{2Y}}{1 + (k-1)\, r_{2Y}}$$

in which r_{2x} is the estimated score reliability of the observations with one rater, r_{2Y} is the obtained inter-rater reliability from the observations with two raters, and k = .5, the percent reduction in the number of raters.

have *average* scores for the six major components that appear relatively low, and much below the "4" rating of "very descriptive." However, these average scores on the six components of the protocol were still much higher than the non-constructivist lessons. Therefore, when examining the component scores of the protocol, it is important to recognize that these are *averages* of the criteria ratings, and that seldom are the component scores as high as the overall ratings for high constructivist lessons.

A second scoring characteristic that became evident was that teacher lecture and discussion approaches to teaching could still provide relatively high scores on the constructivist scale, even though student active participation, group work, and projects are many times seen as an important component of what is thought of as constructivist teaching. And, conversely, simply because a teacher attempted to use group work or project-based learning, it was not a guarantee of a high observation rating, either on the protocol components or the holistic rating. While only about 17% of the lessons were considered highly "constructivist" overall, the two components receiving the highest ratings related to depth of understanding (component 6) and conceptual development (component 1). One possible explanation for this finding is that observers did see a number of challenging and effective "direct instruction" lessons. The following contrast in lessons illustrates this scoring characteristic of the protocol.

One of the lessons observed was high school history and was primarily a lecture. The teacher used PowerPoint (including video clips and pictures) and an outline to engage the students. He made connections between historical and current events (Pearl Harbor and September 11), while students commented on similarities between Kamikaze pilots and the terrorist bombers, and made distinctions between war and terrorism. Although the lesson utilized lecture techniques, students were engaged, communicative, and interested in the presentation, and the content was challenging and involved a real world context. The lesson was scored relatively high on conceptual understanding, creating personal meaning, real world contexts, and challenging curriculum, and scored low only on active participation, exploration and research.

In contrast, in a third grade classroom a teacher attempted to use a number of strategies generally related to constructivist teaching for finding perimeter. The teacher attempted to lead students to understand the idea of perimeter, starting the lesson with definitions (length, width, measure, perimeter). She then asked for examples of "where you might want to know perimeter." Students had no ideas, so she asked, "What might you want to know the distance around? [no response] . . . for your pet?" It was clear that she wanted them to come up with the example of fencing for a yard. This was lost on the students, however, who instead suggested finding the perimeter of a dog's neck, dog's food, or a doghouse. When it was clear that these examples would not work, she used the fence example and showed the actual calculation. This was followed by an assignment to give a rule for finding the perimeter of a given square and then to actually calculate the value. Students were unclear about this, however, and most instead just solved the problem (3+3+3+3=12). In the end, her examples, questions, and explanations ("the formula is length plus width times length *times* width") proved very confusing for the students (as well as the researchers), and the remainder of the class period was spent answering questions and trying to re-teach the concept. In this instance, even though the teacher attempted to use several important constructivist principles, the lesson was scored low on most components of the protocol.

These two examples reflect one of the characteristics of the protocol scoring. *Attempting* to use constructivist principles does not insure "powerful teaching," and therefore does not insure a high score on the protocol. Conversely, the use of more "traditional" instructional techniques does not preclude the lesson from receiving a high score if a number of the constructivist elements are present. In summary, observers concluded that direct instruction was often more powerful, engaging, and instructive than attempts at group work, student projects and other constructivist approaches done poorly. Therefore, some lessons employing more traditional techniques received higher scores than did some lessons attempting to use constructivist principles.

Sample Constructivist and Non-Constructivist Lessons and Scoring

Five vignettes in Appendix F are examples of "typical" lessons and their ratings using the observation protocol. The first vignette summarizes a lesson given a "4" holistic rating, indicating that the observer rated the overall lesson as very constructivist in nature and the highest rating possible according to the scale employed for this study. An explanation of the scoring for each the six components and the average score for each component is also provided. Each of the other vignettes follow this pattern, with two vignettes provided for a holistic scored "1" lesson, typifying a relatively low level experience for students, lacking depth and appropriate or relevant learning strategies.

Total Sample Scores and Frequencies on the TAOP

The means and standard deviations for the 27 indicators of powerful teaching and learning used on the TAOP for the entire sample of 669 classrooms are shown in Table 5. The means and standard deviations for the 7 lesson components and holistic rating of the TAOP for the 669 classrooms are shown in Table 6. The seven component and holistic rating scores were rounded to the nearest whole number and the frequency of these rounded scores for the 7 lesson components and the holistic score are shown in Figures 2 through 9.

TABLE 5. RANK ORDER BY MEANS OF THE 27 INDICATORS FOR OBSERVATIONS IN 669 CLASSROOMS.

TAOP Indicator #	TAOP Item/Indicator	N	Mean	Std. Deviation
1.	Students use appropriate methods and tools of the subject area to acquire and represent information	647	2.8	1.1
27.	Assessment criteria focus on demonstration of knowledge and conceptual understanding of core concepts.	49	2.7	1.2
6.	Students use an appropriate learning strategy to gain meaning	647	2.5	1.3
	Overall conclusion: How constructivist was this lesson?	669	2.4	1.0
2.	Students develop conceptual understanding	647	2.3	1.3
26.	Assessment requires students to communicate learning through elaborated writing, speaking, modeling, diagramming, or demonstrating.	49	2.3	1.3
24.	Lesson presented emphasizes conceptual understanding, not just recall or superficial understanding.	647	2.3	1.3
25.	Central ideas and concepts of the subject are covered in depth.	647	2.2	1.3
4.	Students construct knowledge by manipulating information and ideas to solve complex problems, discover new meaning, and/or develop understanding	647	2.1	1.4
10.	Teacher provides focused feedback and questions to students which probe students' conceptual understanding and lead to sense making.	647	1.9	1.4
5.	Students communicate conceptual understanding through elaborated writing, speaking, modeling, diagramming, or demonstrating	647	1.8	1.5
3.	Students demonstrate thinking by using vocabulary and fundamental concepts of subject area	647	1.7	1.3
18.	Students generate their own ideas, questions, or hypotheses.	647	1.7	1.5
7.	Students rethink (revise) work based on data, self-evaluation and/or constructive feedback from peers/teacher	647	1.6	1.4
11.	Students and/or students and teacher engage in substantive conversation which builds knowledge and develops critical thinking.	647	1.5	1.3
17.	Students work collaboratively to share knowledge, complete projects, and/or critique their work.	647	1.4	1.5
13.	Teacher or student connects knowledge to relevant personal experiences.	647	1.4	1.5
8.	Students consider alternatives and/or multiple ways to investigate and problem solve	647	1.3	1.4
21.	Teacher activates and accesses prior knowledge of students.	647	1.3	1.3
13.	Teacher or student connects knowledge within or across disciplines or to a real world problem.	647	1.3	1.6

9.	Students intentionally reflect on their own learning (metacognition).	647	1.0	1.3
23.	Lesson builds on diverse cultural traditions, student interests and experiences.	647	.9	1.4
22.	Student needs and strengths are accommodated through differentiated learning.	647	.8	1.3
14.	Instruction uses community resources or data.	647	.5	1.2
19.	Students plan and/or carry out independent research.	647	.4	1.1
16.	Students interact with the world outside the school via field-based experiences or technology.	647	.4	1.0
20.	Students independently access/use print media, equipment, or technology.	647	.4	1.0
15.	Students produce a product or performance for an audience beyond the classroom.	647	.2	.9

TABLE 6. RANK ORDER BY MEANS OF THE SEVEN LESSON COMPONENTS OF THE TAOP FOR OBSERVATIONS IN 669 CLASSROOMS AND THE HOLISTIC RATING SCORE.

Component #	Lesson Component	N	Mean	Std. Deviation
7	Summative Assessment allows student to exhibit higher order thinking and construct knowledge.	52	2.5	1.2
6	Students are presented with a challenging curriculum designed to develop depth of understanding.	647	2.2	1.3
1	Student work shows evidence of conceptual understanding, not just recall.	647	2.1	1.1
2	Students are engaged in activities to develop understanding and create personal meaning through reflection.	647	1.6	1.1
4	Students are engaged in active participation, exploration and research.	647	1.0	1.0
5	Teacher uses diverse experiences of students to build effective learning.	647	1.0	1.o
3	Apply knowledge in real world contexts.	647	.7	.9
Holistic Rating	Overall Conclusion: How constructivist was this lesson?	669	2.4	1.0

Figure 2. Frequencies of Scores for Lesson Component 1 for 647 Classroom Observations

Student Work Shows Evidence of Conceptual Understanding, Not Just Recall.

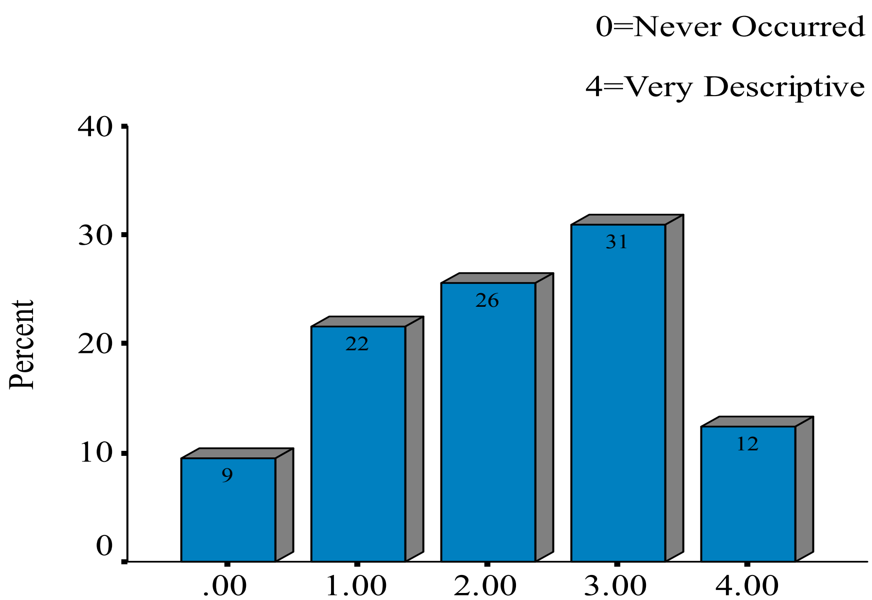


Figure 3. Frequencies of Scores for Lesson Component 2 for 647 Classroom Observations

Students Are Engaged in Activities to Develop Understanding and Create Personal Meaning Through Reflection.

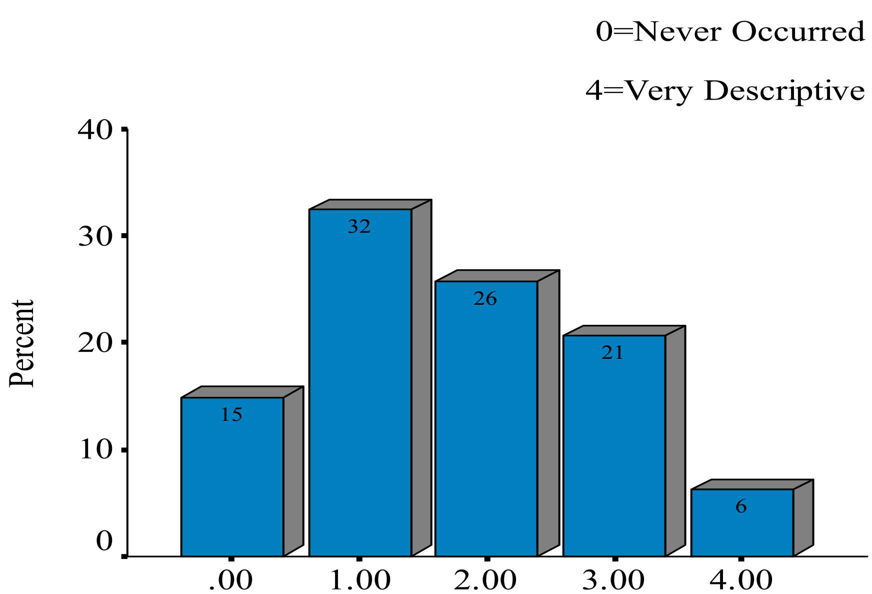


Figure 4. Frequencies of Scores for Lesson Component 3 for 647 Classroom Observations

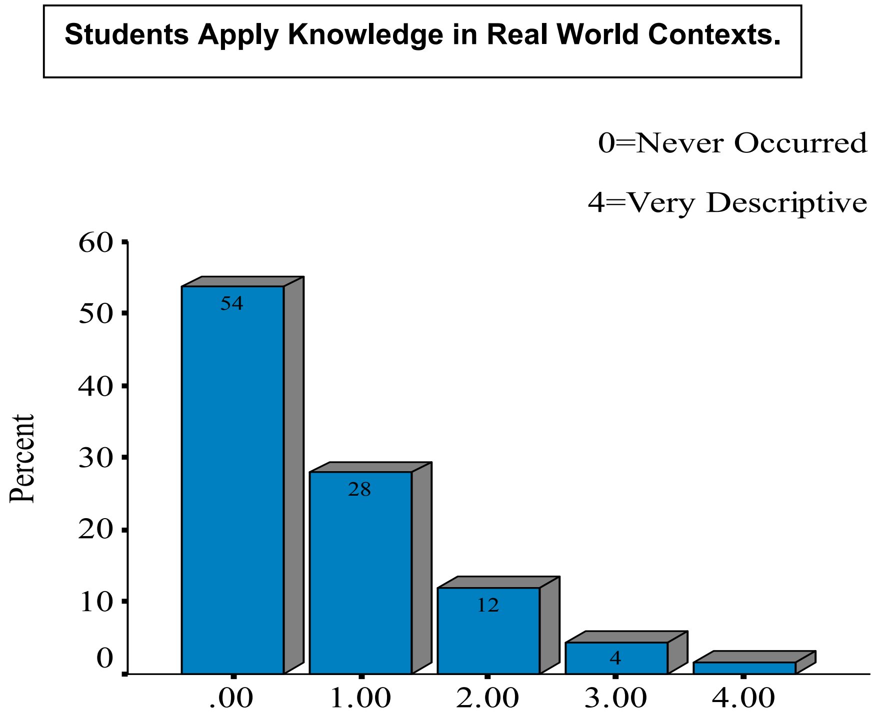


Figure 5. Frequencies of Scores for Lesson Component 4 for 647 Classroom Observations

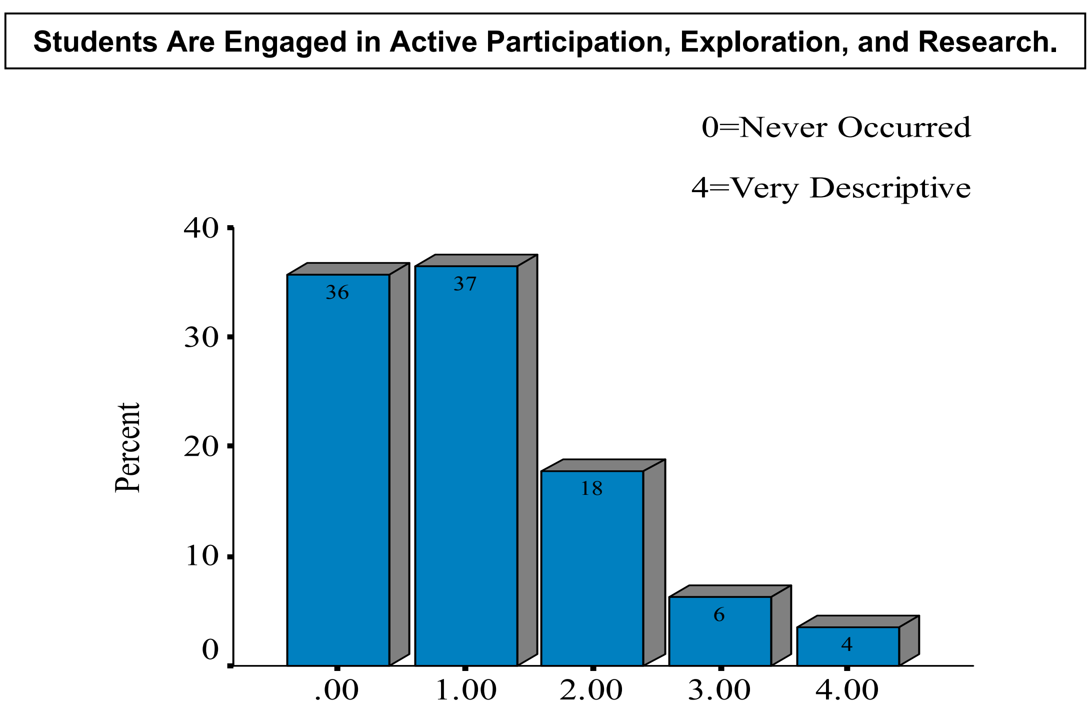


Figure 6. Frequencies of Scores for Lesson Component 5 for 647 Classroom Observations

Teachers Use Diverse Experiences of Students to Build Effective Learning.

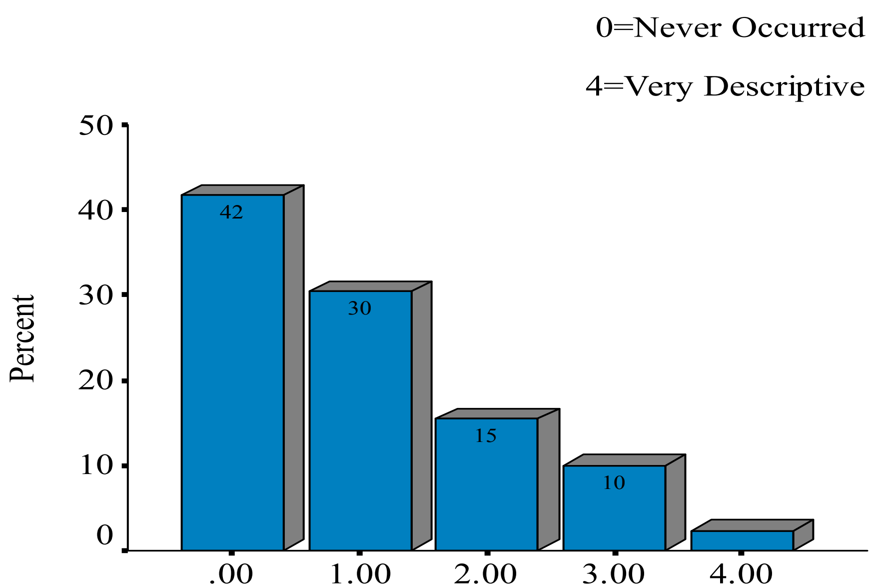


Figure 7. Frequencies of Scores for Lesson Component 6 for 647 Classroom Observations

Students Are Presented with A Challenging Curriculum Designed to Develop Depth of Understanding.

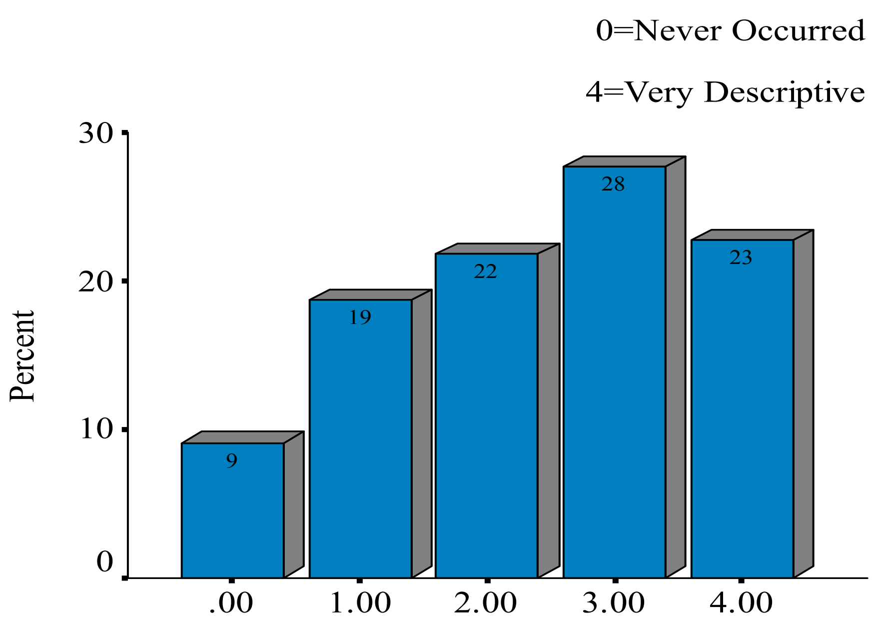


Figure 8. Frequencies of Scores for Lesson Component 7 for 52 Classroom Observations

Summative Assessment Allows Student To Exhibit Higher Order Thinking and Construct Knowledge.

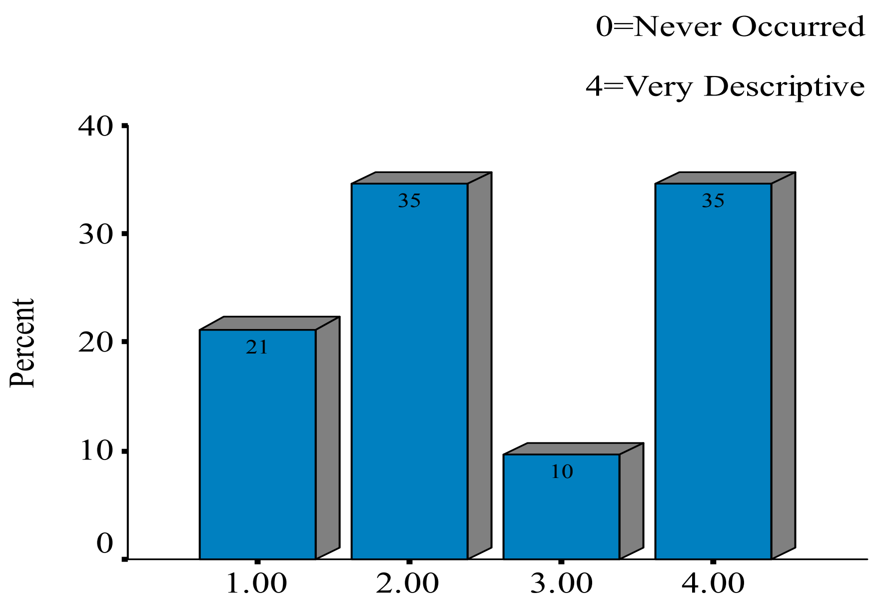
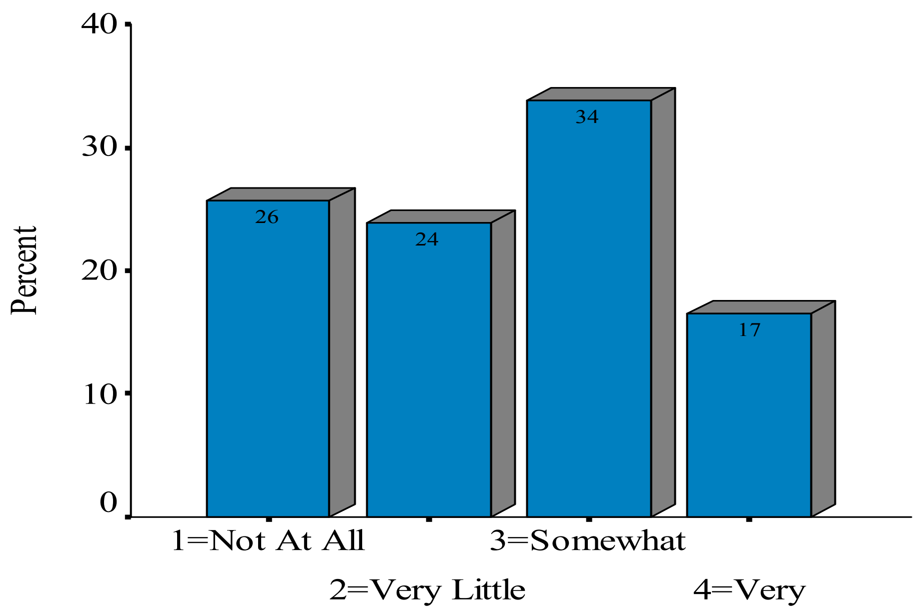


Figure 9. Frequencies of Scores for Holistic Rating for 669 Classroom Observations

Overall, How Constructivist Was This Lesson?



The data presented in Figure 9 show that in only about 17% of the classrooms were teachers engaging students in lessons that the observers could label as “very constructivist.” The observers labeled the lessons as “somewhat constructivist” in an additional 1/3 of the classes. In the remaining 50% of the classes the observers saw little or few constructivist principles being used.

Of the lessons observed the highest component rating was for the 52 instances when summative assessments were being employed (Table 5). In about one half of these assessments students were asked to exhibit higher order thinking and to construct knowledge (Figure 8). The data in Figures 2 and 7 show that students are being presented with a challenging curriculum in about one-half of their lessons, but in less than one half of their lessons did their work, if there was any, show evidence of conceptual understanding. The constructivist principles reflected in components 3 (Students Apply Knowledge in Real World Contexts), 4 (Students are Engaged in Active Participation, Exploration, and Research) and 5 (Teacher Use Diverse Experiences of Students to Build Effective Learning) were used only about 12% of the time or less during the lessons.

Pearson *r* intercorrelations among the 7 lesson component scores and the holistic rating scores of the TAOP are shown in Table 7. As expected, all 7 components correlate significantly with each other and with the holistic score. The strength of these correlations suggests that the instrument reflects internal consistency with a major theoretical construct underlying the 7 components and the holistic score.

TABLE 7. INTERCORRELATIONS OF THE SEVEN LESSON COMPONENT SCORES AND HOLISITIC RATING SCORE OF THE TAOP.

TAOP Component	1	2	3	4	5	6	7	Holistic Rating
1								
2	.86**							
3	.50**	.52**						
4	.67**	.73**	.60**					
5	.49**	.60**	.65**	.55**				
6	.85**	.79**	.48**	.61**	.46**			
7	.76**	.68**	.72**	.75**	.46*	.64**		
Holistic Rating	.85**	.81**	.56**	.70**	.51**	.85**	.89**	

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

TAOP Scores for Elementary, Middle/Junior High, High, Alternative and Voc-tech Classrooms and Schools

Means and standard deviations on the TAOP were calculated for the entire sample of elementary classrooms (n=200), middle/junior high classrooms (n=136), high school classrooms (n=268), alternative school classrooms (n=22), and voc-tech classrooms (n=21). These data are shown in Table 8. The data were also aggregated at the school level, and those means and standard deviations by classroom type are shown in Table 9. Both methods of calculations produced very similar mean scores by school type. However, the standard deviations in Tables 8 and 9 are noticeably different. The mean standard deviation for the 669 classrooms is 1.07, and for the data aggregated by school the mean is .62. The variance among the 669 classrooms is more than three times greater than among the 34 schools, and for the holistic score the variance is 4 times as large[9]. These data indicate that there is considerably more variability among the *classrooms* than among the schools.

TABLE 8. TAOP COMPONENT AND HOLISTICS RATING SCORES BY TYPE OF SCHOOL.

School Group		Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Avg. of Comp 1-6	Holistic Rating
Elementary	m	2.06	1.61	.56	.86	1.07	2.13	2.28	1.38	2.43
	sd	1.05	1.02	.68	.77	1.02	1.28	1.10	.80	1.01
	n	200	200	200	200	200	200	5	199	202
Middle/ jr.high	m	2.01	1.53	.71	1.00	1.02	2.01	2.29	1.37	2.21
	sd	1.19	1.09	.91	1.10	.99	1.26	1.23	.94	1.06
	n	136	136	136	136	136	136	16	135	144
High school	m	2.14	1.56	.65	.87	.81	2.27	2.42	1.38	2.33
	sd	1.13	1.05	.82	.93	.88	1.22	1.10	.83	1.01
	n	268	268	268	268	268	268	26	265	280
Alternative	m	2.86	2.47	1.75	1.74	2.43	2.85		2.35	3.09
	sd	.94	1.12	1.05	1.20	1.35	.89		.93	1.02
	n	22	22	22	22	22	22		22	22
Voctech	m	3.12	2.56	2.66	1.90	1.76	2.93	3.30	2.49	3.24
	sd	.84	.89	1.05	1.18	1.35	1.21	1.30	.84	.77
	n	21	21	21	21	21	21	5	21	21
Total	m	2.15	1.63	.74	.96	1.02	2.21	2.45	1.45	2.39
	sd	1.12	1.07	.91	.97	1.03	1.25	1.16	.88	1.04
	n	647	647	647	647	647	647	52	642	669

[9] Because a standard deviation is a non-interval statistic, direct ratio comparisons are inappropriate. The values must be squared to obtain the variance value, and it is those variance values that may then be compared as ratios. In this example, the standard deviation of 1.07 is squared to obtain a variance value of 1.14 and the standard deviation of .62 is squared to obtain a variance of .38. These variance values may then be compared in a ratio format.

TABLE 9. TAOP COMPONENT AND HOLISTIC RATING SCORES AGGREGATED AT THE SCHOOL LEVEL.

School Group		Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Avg. of Comp 1-6	Holistic Rating
Elementary	m	2.06	1.59	.52	.83	1.03	2.11	2.04	1.35	2.39
	sd	.49	.45	.23	.30	.46	.46	.92	.35	.42
	n	15	15	15	15	15	15	4	15	15
Middle/ Jr. high	m	2.08	1.61	.75	1.07	1.04	2.09	2.44	1.43	2.28
	sd	.71	.68	.45	.57	.45	.80	.93	.56	.61
	n	8	8	8	8	8	8	6	8	8
High schools	m	2.15	1.56	.64	.86	.80	2.26	2.53	1.38	2.33
	sd	.26	.20	.13	.21	.23	.26	.79	.20	.15
	n	7	7	7	7	7	7	7	7	7
alternative	m	2.86	2.47	1.75	1.74	2.43	2.85		2.35	3.09
	sd	.78	.53	.40	.37	.44	.69		.54	.51
	n	2	2	2	2	2	2		2	2
Voctech	m	3.32	2.80	2.81	2.00	1.90	3.15	3.13	2.66	3.37
	sd	.64	.80	.50	.30	.46	.73	1.24	.57	.42
	n	2	2	2	2	2	2	2	2	2
Total	m	2.20	1.71	.81	1.01	1.12	2.24	2.46	1.51	2.45
	sd	.61	.58	.65	.48	.57	.60	.87	.53	.51
	n	34	34	34	34	34	34	19	34	34

The data from the 669 classroom observations were analyzed using the SPSS General Linear Model (Multivariate) to determine if significant differences existed among the types of classrooms observed. Complete results of these analyses are provided in Appendix G. The statistical analysis shows that the classrooms in the elementary, middle/junior, and high schools were all quite similar in the scores they received on the TAOP, and the classrooms in the alternative and voc-tech schools were quite similar in scores. However, the alternative/voc-tech classrooms scored significantly higher on most of the components, including the holistic rating, than did the classrooms in the elementary, middle/junior, and high schools.

TAOP Scores by Subject Matter

Means and standard deviations on the TAOP were calculated for groups of classrooms based on the subject matter of the lesson. These data are shown in Table 10. The data were analyzed using the SPSS General Linear Model (Multivariate) to determine if significant differences existed among the types of classrooms based on the subject matter of the class. Complete results of these analyses are provided in Appendix H. The statistical analysis shows that the classrooms categorized as “integrated” scored significantly higher on the TAOP than were the traditional subject matter classes. Of the four traditional subject matter lessons, science and English classes showed some tendency to be scored higher on the TAOP, and math classes received the lowest scores. However, the differences among these four subject matter area classes were not large.

TABLE 10. TAOP COMPONENT AND HOLISTICS RATING SCORES AGGREGATED BY CLASS SUBJECT MATTER

Subject		Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Avg. of Comp 1-6	Holistic Rating
English	m	2.14	1.69	.82	.97	1.29	2.14	2.30	1.51	2.42
	sd	1.10	1.07	.88	.93	1.11	1.29	1.48	.90	1.03
	n	203	203	203	203	203	203	10	202	204
Math	m	1.93	1.52	.24	.68	.73	2.09	2.25	1.20	2.16
	sd	1.06	1.00	.49	.77	.79	1.21	.87	.74	1.00
	n	166	166	166	166	166	166	14	165	174
Science	m	2.36	1.61	.72	1.09	.81	2.44	2.86	1.50	2.56
	sd	1.15	1.11	.89	.99	.95	1.23	1.10	.87	1.05
	n	113	113	113	113	113	113	11	112	117
Social Studies	m	2.05	1.49	.89	1.00	.91	2.16	2.09	1.40	2.32
	sd	1.17	1.10	.87	1.10	.96	1.25	1.07	.92	1.04
	n	125	125	125	125	125	125	12	123	134
Integrated or Other	m	2.78	2.34	1.95	1.59	1.77	2.64	3.30	2.18	2.99
	sd	.99	.93	1.20	1.07	1.26	1.16	1.30	.87	.87
	n	40	40	40	40	40	40	5	40	40
Total	m	2.15	1.63	.74	.96	1.02	2.21	2.45	1.45	2.39
	sd	1.12	1.07	.91	.97	1.03	1.25	1.16	.88	1.04
	n	647	647	647	647	647	647	52	642	669

Relationship of the TAOP and Teacher Perspectives Questionnaire

During the 2000-2001 school year, teachers in over 200 Gates grantee schools completed the *Teacher Perspectives Questionnaire* (TPQ) as part of baseline assessment evaluation activities (see Fouts, et al., 2001a; 2001b). Factor analyses of the questionnaire resulted in a Constructivist Teaching Scale, measuring the degree to which teachers believe that constructivist teaching methods are used regularly at the school. The schools selected for this observation study were among those schools that had completed the questionnaire during the previous school year. In each of the 34 schools at the time of the observation study one year later a second administration of a shortened version of the *Teacher Perspectives Questionnaire* containing only the constructivist Teaching Scale took place with the sample of teachers whose classrooms were observed. Valid returns on this second administration were received from 30 of the 34 schools. Thus, for 30 schools there are three available sets of scores: Constructivist Teaching Scale score from administration #1 of the TPQ; Constructivist Teaching Scale score from administration #2 of the TPQ; and the TAOP scores from the classroom observations. For 34 schools there are two sets of scores: Constructivist Teaching Scale score from administration #1 of the TPQ, and the TAOP scores from the classroom observations.

Correlations among these sets of scores for the 30 schools are shown in Table 11. Most notable is the very strong correlation, .92, between the Constructivist Teaching Scale score from the first administration of the TPQ and the score from the second administration one year later. Three of the four correlations between the two classroom observation scores and the Constructivist Teaching Scale of the TPQ are statistically significant. These correlations are in the low to moderate range for this type of study. However, the calculations are based on data aggregated at the school level for which there is somewhat limited variability among schools, suggesting that these correlations may underestimate the strength of the relationship that may exist among these variables. When a statistical procedure is used to correct for this restriction of range problem, the correlations between the observation results and the TPQ increase to as high as .72[10]

TABLE 11. INTERCORRELATIONS OF THE CONSTRUCTIVIST TEACHING SCALE FROM THE TEACHER PERSPECTIVES QUESTIONNAIRE AND CLASSROOM OBSERVATION RESULTS.

Score	Administration #1 of Teacher Questionnaire	Administration #2 of Teacher Questionnaire	Classroom Observation Holistic Scoree	Classroom Observation Avg. of Comp. 1-6
Administration #1 of Teacher Questionnaire				
Administration #2 of Teacher Questionnaire	.92** (n=30)			
Classroom Observation Holistic Score	.40* (n=34)	.33 (n=30)		
Classroom Observation Avg. of Components 1-6	.35* (n=34)	.37* (n=30)	.91** (n=34)	

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

[10] In educational research using classroom scores aggregated at the school level, it common for the variance among classrooms to be larger than the variance among schools. The standard deviation values in Tables 8 and 9 and the corresponding discussion shows that this is indeed the case in this study. There is greater variability in classroom practices as measured by the TAOP than demonstrated by the variation among schools, thus creating a restriction of range of scores on that variable. Statistical procedures have been developed to estimate the strength of the correlation if restriction of range was not a problem. The standard formula is:

$$\sim r_{xy} = \frac{r_{t\,(xy)}\ [S_x / S_{t(x)}]}{\sqrt{1 + r^2_{t(xy)}[\ S^2_x / S^2_{t(x)}] - r^2_{t(xy)}}}$$

in which
$\sim r_{xy}$ is the adjusted correlation
$r_{t\,(xy)}$ is the correlation between X and Y with the range of X truncated
S_x is the unrestricted standard deviation of X (estimated from prior data or from some knowledge of the population distribution
$S_{t(x)}$ is the truncated standard deviation of X
If the standard deviation of 1.04 for the holistic observation score from the 669 classrooms is used in this formula as the unrestricted standard deviation, the adjusted correlation to administration #1 of the TPQ is .72. In other words, if the schools showed as much variability as the individual classrooms within those schools do, the correlation would be much stronger. At any rate, the resulting significant positive correlations between the observation results and teacher perceptions on the TQP does provide evidence of validity to the Constructivist Teaching Scale on the TPQ.

SUMMARY AND DISCUSSION

This study is part of the on-going program evaluation of the Bill & Melinda Gates Foundation's Model Schools Initiative and Model Districts Initiative in the state of Washington. In developing the Teaching Attributes Observation Protocol (TAOP) a conceptual framework was identified based on extensive literature on constructivist teaching. From this framework and from the foundation's written materials we identified important components and indicators of constructivist teaching and implications for the classroom. We then produced an observation protocol with 7 lesson components and a number of indicators under each component. The content validity of the instrument was then checked against the literature and existing observation instruments.

Following an extensive training period, classroom observations were conducted in 669 classrooms in 34 schools over a four month period of time. Provisions were made for continual checks for inter-rater reliability and agreement, and the results suggest that there was a high degree of consistency in the rating process. Measuring an abstract construct like constructivist teaching required a continual discussion among observers and clarification throughout the process.

The general findings of this study are that strong constructivist teaching was observable in about 17% of the classroom lessons. The other 83% of the lessons observed may have contained some elements of constructivist teaching, but as many as one-half of the lessons observed had very little or no elements of constructivist teaching present. More constructivist teaching appeared to take place in alternative schools and in integrated subject matter classes than in traditional schools or subjects. There appeared to be no differences among the elementary, middle/junior high and high schools as to the degree to which constructivist practices were used. Finally, the results of the classroom observations do suggest that there is some validity to the Constructivist Teaching Scale of the *Teacher Perspectives Questionnaire.*

The opportunity to observe 669 classroom lessons in total, and over 200 lessons in some cases for an individual observer, is an experience that few educators or researchers have had. At the conclusion of the four months of observations, the four observers were gathered together and asked to reflect on their experiences. Their informal observations about the schools and classes, while not research findings in the formal sense, are intriguing and appropriate for this discussion section. There was general consensus among the four researchers in the following areas.

First, it was apparent that alternative secondary and technical schools provided a constructivist learning environment for students to a greater degree than did other schools. This was particularly true in the way teachers connected students' school experiences to the "real world." Although this may be inherent in an alternative curriculum (dental assisting and welding, for instance, may be more easily connected with the outside world than calculus or chemistry), alternative secondary school teachers in all subject areas seemed to be more intentional in relating knowledge and skills to application. Moreover, students appeared to have greater ownership in their work. For example, students in a welding class were assigned roles similar to those found in a real shop (foreman, etc). They took these roles seriously and offered a tour of the department during which they discussed the various projects underway. They were enthusiastic and eager to share what they learned in class, including the properties of different welding materials, economic considerations of ordering supplies, recent developments in welding equipment, and the importance of customer satisfaction. These students enjoyed their work in class and appeared to believe that it was important and "real." In addition to a strong academic program, observers noted a mutual respect between teachers and students, and among students themselves. While teachers were obviously "in charge," they appeared to regard students as responsible, interested, and committed learners.

Second, curricular materials, both elementary and secondary levels, were often thoughtful and included impressive critical-thinking elements. In some cases, the curriculum was actually superior to the actual teacher instruction.

Third, behavior problems were relatively few, and in only a handful of schools did discipline appear to be a serious issue. In those cases, however, it defined the culture of the school. In some schools, for example, physical and verbal disrespect among students was evident not only in classrooms, but also in the hallways and common areas. Likewise, the degree of student disrespect towards teachers was surprising in its frequency and intensity.

Fourth, while few in number, there were instances where researchers got an immediate sense of a school culture. In some, the culture was one of academics. While all elementary schools would probably report that they have focused their efforts on literacy, this was not readily apparent to observers even after being in a building for an entire day. In one school, however, it was immediately clear that the school not only "talked" literacy, but had made it the focus of the entire building. Staff, parents, and students were all involved, and there was no missing the direction of the school. In other schools, one could feel a shared sense of purpose and collaboration, not only in the classroom, but throughout the school. Interactions between students and teachers were respectful, and expectations for learning seemed to be higher.

Fifth, a related, but larger issue was the amount of time spent in elementary schools on "housekeeping" and managerial tasks, travel time, special activities, and the like. More than once, researchers found themselves moving from one classroom to another in an attempt to find an academic lesson (language arts, math, science, or social studies) to observe. Whether students were participating in music, P.E., library, computer lab, band, recess, class birthday parties, rehearsal for a school program or assembly, there seemed to be an excessive amount of time devoted to non-academic tasks. One teacher recognized this dilemma in noting that, "…five of the students are gone to band for the next 30 minutes. We don't teach when the kids are at band." In this case, students were involved in independent activities or "catch up" work.

Sixth, the wide range of expectations teachers had of their students intrigued the researchers. For example, in one case high school biology students were given a copy of an animal cell and asked to color it and then cut and paste paper labels to each part. A culminating activity for a junior high social studies unit involved drawing and coloring one of Columbus's ships. At the other end of the spectrum, students in a high school science class were involved in isolating and observing DNA from a sample of wheat germ, while a class of elementary students participated in a book discussion that centered on frank and powerful themes related to real-life situations.

Finally, observers were not surprised at the extensive use of cooperative groups. However, they *were* surprised at the number of instances where it was used ineffectively. While students were frequently asked to complete assignments and projects in groups, this was often done with little or no direction from the teacher in terms of goals or expectations. As a result, students approached these tasks with little sense of purpose, and true collaboration was minimal. Given all the time and training educators have invested in the use of cooperative learning strategies over the past 20 years, it was disconcerting to find it used ineffectively in so many cases.

These last two topics are of particular interest and pertain to other research in the area of constructivist teaching. One of the features of the TAOP that became evident throughout this study is that scoring high as a constructivist lesson is less dependent on specific teaching strategies and more dependent on certain types of intellectual demands placed on the student. While this was not necessarily our intent going into this project, it became evident as the construct of constructivism emerged from the literature. In fact, our findings and the nature of the TAOP are in accord with the work of Newmann, et. al. and the Consortium on Chicago School Research (Newmann, Lopez, & Bryk, 1998; Bryk, Nagaoka, & Newmann, 2000; Newmann, Bryk, & Nagaoka, 2001). Their research has demonstrated that it is the quality of the intellectual work that students undertake that makes the difference. They use the terms "authentic intellectual work," but this phrase incorporates many of the ideas that are the basis of "constructivist teaching" used in our study. An important finding from their research is worth noting here. In their examination of the relationship of student work and student test scores they found that no teaching strategy ensured that the student would face "high quality intellectual demands." They found numerous examples of "hands-on" or "active-learning" classroom projects that provided little, if any opportunity for intellectual growth. On the other hand, they did find "demanding 'teacher-centered' lecture and question-and-answer instruction that requires students to think deeply about issues important in their lives." They go on to say:

> 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 (Newmann, Bryk, & Nagaoka, 2001, p 31).

We believe this is the case with the TAOP as well, and this has been mentioned throughout this report. Certain teaching strategies might increase the score for the observation to some degree, but without intellectually demanding activity (authentic learning)in the lesson, the scores were relatively low.

During the 2002-2003 school year we will continue our work in this area by conducting additional classroom observations in 16 high schools in the state of Washington and by adding a study of the intellectual quality of student work using the model developed by Newmann, et al.

REFERENCES

Accelerated Schools Project (2001). *Accelerated Schools Project tools for assessing school progress. Observation notes.* Storrs, Ct: Author. Available at: http://www.acceleratedschools.net/main_acc.htm.

American Institutes of Research/SRI International (2002). *Targeted literature review of major constructs and their* components*: Evaluating the National School District and Networks Grants Program.* Unpublished manuscript.

Arizona Collaborative for Excellence in the Preparation of Teachers (2000). *Reformed Teaching Observation Protocol Reference Manual.* (Technical Report No. IN00-3). Arizona State University.

Bransford, J.D**.** & Vye, N. J. (1989). A perspective on cognitive research and its implication for research. In L. Resnick & L. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research*. Alexandra, VA: ASCD.

Brooks, J. G. & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: ASCD.

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

Caine, R. N. & Caine, G. (1991). *Making connections: Teaching and the human brain.* Alexandria, VA: ASCD.

Caine, R. N. & Caine, G. (1997). *Education on the edge of possibility.* Alexandria, VA: ASCD.

CETP Core Evaluation (2001). *Classroom Observation Handbook*. College of Education & Human Development, University of Minnesota. Paper presented at the American Educational Research Association, Seattle, WA, April 2001.

Chaille, C. & Britain, L. (1997). *The young child as scientist*. (2nd ed) New York: Longman

Duffy, T.M, & Cunningham, D.J. (1996). Constructivism: Implications for the design and delivery of instruction. In D.H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp.170-198). New York: Macmillan.

Educational Testing Service (2000). PATHWISE Classroom Observation System Orientation Guide. ETS: Author.

Fan, X, & Chen, M. (2000). *Published* studies of interrater reliability often overestimate reliability: Computing the correct coefficient. *Educational & Psychological Measurement, 60,* 532-541.

Fosnot, C. T. (1996). *Constructivism theory, perspectives, and practice.* New York: Teachers College Press.

Fouts, J.T. (2000). *Research on computers and education: Past, present and future.* Seattle, WA: Bill & Melinda Gates Foundation. Available at http://www.gatesfoundation.org/education/evaluation/default.htm

Fouts, J.T., Baker, D.B., Riley, S.C., Abbott, M.L., & Robinson, H.L. (2001a). *The Bill & Melinda Gates Foundation's Model District Initiative: Year 1 evaluation results.* Seattle, WA: Fouts & Associates and the Washington School Research Center. Available at http://www.gatesfoundation.org/education/evaluation/default.htm

Fouts, J.T., Baker, D.B., Riley, S.C., Abbott, M.L., & Robinson, H.L. (2001b). *The Bill & Melinda Gates Foundation's Model School Initiative: Year 1 evaluation results.* Seattle, WA: Fouts & Associates and the Washington School Research Center. Available at http://www.gatesfoundation.org/education/evaluation/default.htm

Frick, T. & Semmel, M.I. (1978). Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research, 48,* 157-184.

Harrop, A, Foulkes, C. & Daniels, M. (1989). Observer agreement calculations: The role of primary data in reducing obfuscation. *British Journal of Psychology, 80*, 181-189.

Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 67* (2), 227-266.

Lambert, L., Walker, D., Zimmerman, D. P., Cooper, J. E., Lambert, M. D., Gardner, M. E. & Ford Slack, P.J. (1995). *The constructivist leader*. New York: Teachers College.

Means, B. (ed.) (1994). *Technology and education reform. The reality behind the promise..* San Francisco, CA: Jossey-Bass Publishers.

Medley, D.M. & Norton, D.P. (1971). *The concept of reliability as it applies to behavior records.* Paper presented at the meeting of the American Psychological Association, Washington. D.C.

National Council for Social Studies (2001). *Powerful and authentic social studies: PASS.* Washington, D.C.: Author.

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. Washing, 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.

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

Newmann, F. M., Secada, W. G. & Wehlage, G. G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring*. Madison: Wisconsin Center for Education Research.

Resnick, L. B. & Klopfer, L. E. (1989). *Toward the thinking curriculum: Current cognitive research*. Alexandra, VA: ASCD.

Secada, W. & Byrd, L. (1993). *Classroom Observation Scales: School level reform in the teaching of mathematics*. Madison, WI: National Center for Research in Mathematical Sciences Education.

Shaw, D.E. & the President's Committee of Advisors on Science and Technology Panel on Educational Technology (1998). Report to the President on the use of technology to strengthen K-12 education in the United States: Findings related to research and evaluation*. Journal of Science Education and Technology, 7*(2), 115-126.

Sylwester, Robert. (1995). *A celebration of neurons. An educator's guide to the human brain.* Alexandria, VA: ASCD.

Thieman, G. (2000). Factors influencing middle school teachers to change classroom practice in response to standards-based reform. Unpublished doctoral dissertation, Portland State University.

United States Department of Education (1993). *Using technology to support education reform.* Washington, DC: Author. Available: http://www.ed.gov/pubs/EdReformStudies/TechReforms/

Vosniadou, S., DeCorte, E., Glaser,R., & Mandl, H. (Eds.). *International perspectives on the design of technology-supported learning environments.* Mahway, NJ: Lawrence Erlbaum Associates.

Wolf, D., Bixby, J., Glenn, J. III & Gardner, H. (1991). To use their minds well: New forms of student assessment. *Review of Research in Education, 17*, 31-74.

Zemelman, S., Daniels, H. & Hyde, A. (1998). *Best practice: New standards for teaching and learning in America's schools, 2nd ed.* Portsmouth, NH: Heineman

Appendix A The Teaching Attributes Observation Protocol (TAOP)

Teaching Attributes Observation Protocol
Gates Model Districts and Schools Initiatives

November 2001

School:__________________________ Date:________________ Time:___________

Grade:__________________________ Subject:________________________

Teacher (code)___________________ Observer:_______________________

Contextual Background and Activities: In your own words briefly describe the lesson, the classroom setting, classroom environment, resources, content or skills taught, teacher and student activities, student work displayed. If possible, look at assignments, project directions, or assessments in which students are involved during your observation.
(Note: If the teacher gives you a whole unit or project materials spanning several days, focus on that part you are seeing in class during that observation.)

Use the following space for recording Student and Teacher activity during the Observation session, as well as any helpful notes on resources, etc.

Teacher Activity	Student Activity

The next six sections contain items to be rated. Space is provided below each major section for making notes during the observation. After the lesson, use your notes to complete the ratings. Each item should be rated from 0 to 4. Indicate "0" if the item did not occur at all during the lesson. Choose between 1, 2, 3, or 4 depending on whether the item occurred very little, somewhat, quite often, or was very descriptive.

STUDENT WORK SHOWS EVIDENCE OF CONCEPTUAL UNDERSTANDING, NOT JUST RECALL

	Never Occurred				Very Descriptive
1. Students use appropriate methods and tools of the subject area to acquire and represent information. text analysis, creative or expository writing. discussion, oral presentation, reading, interviews. desktop publishing, manipulatives, models, maps, timelines. calculators, primary sources, drawing, graphs, symbols,	0	1	2	3	4
2. Students develop conceptual understanding. organizing information, applying information, considering alternatives, interpreting or evaluating, predicting, comparing, contrasting, analyzing cause & effect, hypothesizing, sequencing, developing a model, simulation, or original creation	0	1	2	3	4
3. Students demonstrate thinking by using vocabulary and fundamental concepts of subject area. literary genres, cause and effect, chemical properties, number theory, probability & statistics	0	1	2	3	4
4. Students construct knowledge by manipulating information and ideas to solve complex problems, discover new meaning, and/or develop understanding. analyzing a story, discussing a public issue using historical evidence or current data to support an opinion analyzing an environmental problem, using symbolic representation theory building where appropriate.	0	1	2	3	4
5. Students communicate conceptual understanding through elaborated writing, speaking, modeling, diagramming or demonstrating. poetry, essays, journals, research papers, letters, response logs, lab reports, dialogue, debate, skit, presentation,	0	1	2	3	4

Comments:

STUDENTS ARE ENGAGED IN ACTIVITIES TO DEVELOP UNDERSTANDING AND CREATE PERSONAL MEANING THROUGH REFLECTION

	Never Occurred				Very Descriptive
6. Students use an appropriate learning strategy to gain meaning. graphic organizer, mapping, drawing pictures outlining, creating a model, journaling discussion, reference to text	0	1	2	3	4
7. Students rethink (revise) work based on data, self-evaluation and/or constructive feedback from peers/teacher.	0	1	2	3	4
8. Students consider alternatives and/or multiple ways to investigate and problem solve.	0	1	2	3	4
9. Students intentionally reflect on their own learning (metacognition). text to self, other texts, world connections; examining own bias or opinion, critique science lab procedures, math reasoning	0	1	2	3	4
10. Teacher provides focused feedback and questions to students that probe students' conceptual understanding and lead to sense making.	0	1	2	3	4
11. Students and/or students and teacher engage in substantive conversation that builds knowledge and develops critical thinking. literature circle, readers' theatre, discuss writing process, simulation, town meeting, debate, generate hypotheses, share and compare results, discuss conclusions, math reasoning	0	1	2	3	4

Comments:

APPLY KNOWLEDGE IN REAL WORLD CONTEXTS

	Never Occurred				Very Descriptive
12. Teacher or student connects knowledge to relevant personal experiences.	0	1	2	3	4
13. Teacher or student connects knowledge within or across disciplines *or* to a real world problem.	0	1	2	3	4
14. Instruction uses community resources or data. guest speakers, materials	0	1	2	3	4
15. Students produce a product or performance for an audience beyond the class. persuasive essay, speech, play, posting student work to a website, letter to the editor, pen pals, brochure, community survey	0	1	2	3	4
16. Students interact with world outside school via field-based experiences or technology.	0	1	2	3	4

Comments:

STUDENTS ARE ENGAGED IN ACTIVE PARTICIPATION, EXPLORATION AND RESEARCH

	Never Occurred				Very Descriptive
17. Students work collaboratively to share knowledge, complete projects, and/or critique their work. writing, response partners, reading groups, research groups, lab groups, math problem solving groups	0	1	2	3	4
18. Students generate their own ideas, questions, or hypotheses.	0	1	2	3	4
19. Students plan and/or carry out independent research. choose research topic, information sources, design lab procedures and search for math patterns	0	1	2	3	4
20. Students independently access/use print media, equipment or technology. books, newspapers, maps, graphs, charts	0	1	2	3	4

Comments:

TEACHER USES DIVERSE EXPERIENCES OF STUDENTS TO BUILD EFFECTIVE LEARNING

	Never Occurred			Very Descriptive	
21. Teacher activates and accesses prior knowledge of students.	0	1	2	3	4
22. Student needs and strengths are accommodated through differentiated learning.	0	1	2	3	4
23. Lesson builds on diverse cultural traditions, student interests and experiences. writing connected to student experience and knowledge, diverse literature, interview family members, lab activities incorporate personal experience, multiple perspectives on numeracy.	0	1	2	3	4

Comments:

STUDENTS ARE PRESENTED WITH A CHALLENGING CURRICULUM DESIGNED TO DEVELOP DEPTH OF UNDERSTANDING

	Never Occurred			Very Descriptive	
24. Lesson presented emphasizes conceptual understanding, not just recall or superficial understanding. comprehension, analysis of literature, support thesis with data, (re)discover theory, math problem solving	0	1	2	3	4
25. Central ideas and concepts of the subject are covered in depth. comprehension , continuity/ change, compare/contrast, cause/effect, number theory, measurement, probability, matter, properties, interdependence	0	1	2	3	4

Comments:

SUMMATIVE ASSESSMENT ALLOWS STUDENTS TO EXHIBIT HIGHER ORDER THINKING AND CONSTRUCT KNOWLEDGE (Choose NA if there was no summative assessment)

	Very Little			Very Descriptive	
26. Assessment requires students to communicate learning through elaborated writing, speaking, modeling, diagramming, or demonstrating.	NA	1	2	3	4
27. Assessment criteria focus on demonstration of knowledge and conceptual understanding of core concepts.	NA	1	2	3	4

Comments:

OVERALL CONCLUSION: HOW CONSTRUCTIVIST WAS THIS LESSON? Circle one answer.

Not at All **Very Little** **Somewhat** **Very**

Appendix B Demographic Characteristics of Schools in the Study

		WA State Average	Pathfinder	COHO/NOMS	Environmental & Adventure School	Hay ES	Madrona ES	Garfield HS	Evergreen HS
School Enrollment (00-01)		**468**	333	567	119	460	363	1733	2031
% F/R Lunch (00-01)		**30.9%**	23.0%	8.8%	0.0%	16.0%	74.0%	20.7%	16.7%
% Asian Enrollment		**6.1%**	7.5%	9.7%	5.0%	11.5%	5.0%	13.4%	7.0%
% Native American Enrollment		**3.7%**	17.7%	6.2%	2.5%	2.2%	4.4%	1.5%	1.0%
% African American Enrollment		**5.4%**	7.8%	5.6%	0.0%	10.9%	70.5%	34.0%	3.8%
% Hispanic Enrollment		**10.5%**	9.0%	7.6%	1.7%	7.8%	8.3%	4.4%	2.4%
Total % Non-white Enrollment		**25.7%**	42.0%	29.1%	9.2%	32.4%	88.2%	53.3%	14.2%
% White Enrollment		**74.3%**	58.0%	70.9%	90.8%	67.6%	11.8%	46.6%	85.7%
Computers per Staff		**NA**	NA	NA	8.80	NA	NA	3.00	1.44
Computers per 10 Students		**NA**	**NA**	**NA**	3.70	NA	NA	1.32	0.88
3rd grade ITBS (00-01) NPR	**Math**	**58.7**	57.5	67.5	NA	72.6	44.4	NA	NA
	Reading	**54.3**	62.9	72.6	NA	71.3	47.1	NA	NA
	Vocabulary	**55.1**	63.5	74.0	NA	71.7	42.6	NA	NA
4th grade WASL (00-01) % passing	**Math**	**42.9%**	50.0%	64.4%	NA	60.8%	4.8%	NA	NA
	Reading	**65.7%**	69.0%	93.3%	NA	75.7%	31.0%	NA	NA
	Writing	**43.0%**	31.0%	51.1%	NA	55.4%	9.5%	NA	NA
6th grade ITBS (00-01) NPR	**Math**	**53.1**	45.9	62.7	68.4	NA	28.0	NA	NA
	Reading	**53.5**	51.3	70.2	76.5	NA	37.4	NA	NA
	Language	**51.2**	44.8	58.1	72.0	NA	35.6	NA	NA
7th grade WASL (00-01) % passing	**Math**	**25.8%**	6.3%	55.8%	65.8%	NA	0.0%	NA	NA
	Reading	**38.7%**	25.0%	61.9%	68.4%	NA	11.6%	NA	NA
	Writing	**46.9%**	12.5%	51.3%	76.3%	NA	20.9%	NA	NA
9th grade ITBS (00-01) NPR	**Math**	**51.1**	NA	NA	NA	NA	NA	67.0	60.0
	Reading	**48.9**	NA	NA	NA	NA	NA	64.1	53.3
	Literature	**51.2**	NA	NA	NA	NA	NA	64.0	56.4
10th grade WASL (00-01) % passing	**Math**	**32.5%**	NA	NA	NA	NA	NA	49.8%	27.9%
	Reading	**57.3%**	NA	NA	NA	NA	NA	52.6%	59.6%
	Writing	**41.2%**	NA	NA	NA	NA	NA	50.2%	31.1%

		WA State Average	Port Angeles HS	Orchards ES	Sacajawea ES	Sealth HS	Happy Valley ES	Washington ES	Linwood ES
School Enrollment (00-01)		**468**	1557	609	267	996	398	477	449
% F/R Lunch (00-01)		**30.9%**	16.8%	49.3%	23.2%	50.8%	28.1%	43.1%	51.2%
% Asian Enrollment		**6.1%**	2.7%	4.9%	14.2%	26.8%	6.3%	2.7%	2.9%
% Native American Enrollment		**3.7%**	5.4%	1.3%	2.2%	2.6%	3.5%	0.2%	5.6%
% African American Enrollment		**5.4%**	1.0%	5.9%	12.7%	17.7%	2.0%	1.9%	4.2%
% Hispanic Enrollment		**10.5%**	1.7%	6.7%	11.2%	21.1%	5.3%	17.2%	1.6%
Total % Non-white Enrollment		**25.7%**	10.8%	18.8%	40.3%	68.2%	17.1%	22.0%	14.3%
% White Enrollment		**74.3%**	89.1%	81.1%	59.6%	31.8%	82.9%	78.0%	85.7%
Computers per Staff		**NA**	1.44	2.51	NA	NA	3.67	1.56	2.68
Computers per 10 Students		**NA**	0.88	1.69	NA	NA	2.19	0.87	1.60
3rd grade ITBS (00-01) NPR	**Math**	**58.7**	NA	47.0	69.2	NA	66.4	50.4	68.4
	Reading	**54.3**	NA	45.0	65.0	NA	59.6	55.8	58.7
	Vocabulary	**55.1**	NA	43.7	62.8	NA	61.9	60.8	51.9
4th grade WASL (00-01) % passing	**Math**	**42.9%**	NA	26.4%	47.2%	NA	75.3%	57.1%	24.2%
	Reading	**65.7%**	NA	59.3%	81.0%	NA	88.3%	76.2%	65.2%
	Writing	**43.0%**	NA	25.3%	39.6%	NA	71.4%	56.0%	31.8%
6th grade ITBS (00-01) NPR	**Math**	**53.1**	NA	NA	NA	NA	NA	NA	63.4
	Reading	**53.5**	NA	NA	NA	NA	NA	NA	53.2
	Language	**51.2**	NA	NA	NA	NA	NA	NA	60.1
7th grade WASL (00-01) % passing	**Math**	**25.8%**	NA	NA	NA	NA	NA	NA	NA
	Reading	**38.7%**	NA	NA	NA	NA	NA	NA	NA
	Writing	**46.9%**	NA	NA	NA	NA	NA	NA	NA
9th grade ITBS (00-01) NPR	**Math**	**51.1**	66.0	NA	NA	49.0	NA	NA	NA
	Reading	**48.9**	59.5	NA	NA	42.2	NA	NA	NA
	Literature	**51.2**	60.6	NA	NA	45.5	NA	NA	NA
10th grade WASL (00-01) % passing	**Math**	**32.5%**	53.7%	NA	NA	14.2%	NA	NA	NA
	Reading	**57.3%**	73.6%	NA	NA	29.0%	NA	NA	NA
	Writing	**41.2%**	52.6%	NA	NA	18.2%	NA	NA	NA

		WA State Average	Glover MS	Prosser Heights ES	Rainier View ES	Havermale Alternative	Schmitz Park ES	Park MS	Lowell ES
School Enrollment (00-01)		**468**	827	472	326	619	321	759	406
% F/R Lunch (00-01)		**30.9%**	60.0%	56.5%	73.9%	51.8%	16.8%	57.8%	8.9%
% Asian Enrollment		**6.1%**	2.1%	0.4%	24.5%	2.1%	10.3%	0.8%	19.2%
% Native American Enrollment		**3.7%**	4.6%	0.2%	1.2%	6.8%	2.8%	0.1%	1.5%
% African American Enrollment		**5.4%**	3.3%	0.8%	55.2%	10.0%	9.7%	1.1%	9.9%
% Hispanic Enrollment		**10.5%**	2.5%	37.7%	9.5%	3.9%	6.9%	32.4%	3.7%
Total % Non-white Enrollment		**25.7%**	12.5%	39.1%	90.4%	22.8%	29.7%	34.4%	34.3%
% White Enrollment		**74.3%**	87.5%	60.8%	9.5%	77.2%	70.4%	65.6%	65.8%
Computers per Staff		**NA**	3.57	4.63	NA	7.14	NA	3.39	NA
Computers per 10 Students		**NA**	2.50	2.90	NA	4.20	NA	2.65	NA
3rd grade ITBS (00-01) NPR	**Math**	**58.7**	NA	47.9	42.7	NA	68.6	NA	97.3
	Reading	**54.3**	NA	45.0	35.8	NA	59.6	NA	94.2
	Vocabulary	**55.1**	NA	43.8	35.5	NA	60.1	NA	93.8
4th grade WASL (00-01) % passing	**Math**	**42.9%**	NA	29.6%	17.0%	NA	53.3%	NA	93.3%
	Reading	**65.7%**	NA	51.5%	31.7%	NA	83.3%	NA	94.4%
	Writing	**43.0%**	NA	40.2%	30.0%	NA	73.3%	NA	87.6%
6th grade ITBS (00-01) NPR	**Math**	**53.1**	NA	NA	NA	NA	NA	38.6	NA
	Reading	**53.5**	NA	NA	NA	NA	NA	36.4	NA
	Language	**51.2**	NA	NA	NA	NA	NA	35.8	NA
7th grade WASL (00-01) % passing	**Math**	**25.8%**	17.9%	NA	NA	NA	NA	14.6%	NA
	Reading	**38.7%**	26.8%	NA	NA	NA	NA	24.0%	NA
	Writing	**46.9%**	43.7%	NA	NA	NA	NA	33.9%	NA
9th grade ITBS (00-01) NPR	**Math**	**51.1**	NA	NA	NA	33.7	NA	NA	NA
	Reading	**48.9**	NA	NA	NA	37.9	NA	NA	NA
	Literature	**51.2**	NA	NA	NA	39.1	NA	NA	NA
10th grade WASL (00-01) % passing	**Math**	**32.5%**	NA	NA	NA	5.8%	NA	NA	NA
	Reading	**57.3%**	NA	NA	NA	26.7%	NA	NA	NA
	Writing	**41.2%**	NA	NA	NA	5.7%	NA	NA	NA

		WA State Average	B.F. Day ES	Aki Kurose MS	Garry MS	Nova	Sunset ES	Wy'East MS	Southgate ES
School Enrollment (00-01)		**468**	284	364	639	235	472	1068	531
% F/R Lunch (00-01)		**30.9%**	47.4%	26.8%	66.2%	0.0%	34.2%	21.0%	18.8%
% Asian Enrollment		**6.1%**	23.2%	6.0%	3.8%	3.0%	5.5%	9.3%	1.3%
% Native American Enrollment		**3.7%**	3.5%	5.0%	4.9%	3.4%	0.6%	0.7%	0.0%
% African American Enrollment		**5.4%**	16.2%	4.4%	5.0%	7.7%	5.7%	3.5%	0.8%
% Hispanic Enrollment		**10.5%**	9.2%	45.3%	1.7%	5.1%	5.9%	4.2%	5.8%
Total % Non-white Enrollment		**25.7%**	52.1%	60.7%	15.4%	19.2%	17.7%	17.7%	7.9%
% White Enrollment		**74.3%**	47.9%	43.7%	84.7%	80.9%	82.2%	82.3%	92.1%
Computers per Staff		**NA**	NA	NA	4.17	NA	NA	3.70	4.29
Computers per 10 Students		**NA**	NA	NA	2.00	NA	NA	1.87	2.26
3rd grade ITBS (00-01) NPR	**Math**	**58.7**	76.6	NA	NA	NA	56.6	NA	59.4
	Reading	**54.3**	65.8	NA	NA	NA	51.5	NA	58.0
	Vocabulary	**55.1**	62.0	NA	NA	NA	51.6	NA	58.4
4th grade WASL (00-01) % passing	**Math**	**42.9%**	33.3%	NA	NA	NA	44.6%	NA	51.4%
	Reading	**65.7%**	57.1%	NA	NA	NA	78.0%	NA	72.4%
	Writing	**43.0%**	45.2%	NA	NA	NA	37.8%	NA	52.4%
6th grade ITBS (00-01) NPR	**Math**	**53.1**	NA	NA	NA	NA	NA	53.5	NA
	Reading	**53.5**	NA	NA	NA	NA	NA	51.9	NA
	Language	**51.2**	NA	NA	NA	NA	NA	51.7	NA
7th grade WASL (00-01) % passing	**Math**	**25.8%**	NA	NA	14.7%	NA	NA	41.8%	NA
	Reading	**38.7%**	NA	NA	25.2%	NA	NA	56.9%	NA
	Writing	**46.9%**	NA	NA	34.2%	NA	NA	71.5%	NA
9th grade ITBS (00-01) NPR	**Math**	**51.1**	NA	40.5	NA	68.5	NA	NA	NA
	Reading	**48.9**	NA	39.7	NA	73.7	NA	NA	NA
	Literature	**51.2**	NA	39.8	NA	73.5	NA	NA	NA
10th grade WASL (00-01) % passing	**Math**	**32.5%**	NA	10.3%	NA	53.7%	NA	NA	NA
	Reading	**57.3%**	NA	33.3%	NA	70.1%	NA	NA	NA
	Writing	**41.2%**	NA	15.0%	NA	49.3%	NA	NA	NA

		WA State Average	Tri City Area Voc.	Mountain View HS	Clark Co. Voc.	Ridge View ES	Southridge HS	Squalicum HS
School Enrollment (00-01)		**468**	12	1930	27	498	1423	1287
% F/R Lunch (00-01)		**30.9%**	65.0%	12.4%	0.0%	19.2%	18.3%	20.3%
% Asian Enrollment		**6.1%**	0.0%	9.4%	3.7%	2.2%	3.0%	6.8%
% Native American Enrollment		**3.7%**	0.0%	0.6%	0.0%	0.4%	0.4%	2.3%
% African American Enrollment		**5.4%**	0.0%	3.8%	0.0%	0.4%	1.8%	1.6%
% Hispanic Enrollment		**10.5%**	8.3%	3.6%	0.0%	16.9%	8.9%	4.8%
Total % Non-white Enrollment		**25.7%**	8.3%	17.4%	3.7%	19.9%	14.1%	15.5%
% White Enrollment		**74.3%**	91.7%	82.6%	96.3%	80.1%	86.0%	84.5%
Computers per Staff		**NA**	3.69	1.10	2.67	2.72	2.00	6.44
Computers per 10 Students		**NA**	1.07	0.60	0.69	1.56	1.07	3.04
3rd grade ITBS (00-01) NPR	**Math**	**58.7**	NA	NA	NA	65.7	NA	NA
	Reading	**54.3**	NA	NA	NA	63.5	NA	NA
	Vocabulary	**55.1**	NA	NA	NA	65.2	NA	NA
4th grade WASL (00-01) % passing	**Math**	**42.9%**	NA	NA	NA	65.5%	NA	NA
	Reading	**65.7%**	NA	NA	NA	83.3%	NA	NA
	Writing	**43.0%**	NA	NA	NA	68.7%	NA	NA
6th grade ITBS (00-01) NPR	**Math**	**53.1**	NA	NA	NA	NA	NA	NA
	Reading	**53.5**	NA	NA	NA	NA	NA	NA
	Language	**51.2**	NA	NA	NA	NA	NA	NA
7th grade WASL (00-01) % passing	**Math**	**25.8%**	NA	NA	NA	NA	NA	NA
	Reading	**38.7%**	NA	NA	NA	NA	NA	NA
	Writing	**46.9%**	NA	NA	NA	NA	NA	NA
9th grade ITBS (00-01) NPR	**Math**	**51.1**	NA	57.3	NA	NA	61.7	58.6
	Reading	**48.9**	NA	55.4	NA	NA	60.1	61.2
	Literature	**51.2**	NA	58.7	NA	NA	63.4	64.0
10th grade WASL (00-01) % passing	**Math**	**32.5%**	NA	29.6%	NA	NA	40.9%	46.2%
	Reading	**57.3%**	NA	72.8%	NA	NA	68.2%	66.8%
	Writing	**41.2%**	NA	45.7%	NA	NA	46.6%	49.3%

Appendix C Letter to Schools

September 25, 2001

Dear Gates Evaluation Coordinator:

As you know, a component of the evaluation activities this year is a classroom observation study in a selection of the schools in the Gates Model District Initiative and the Gates Model School Initiative. We have selected a group of schools from grantees that we believe are representative of the over 250 schools in the state receiving Gates funding. Through this process your school (if you received an individual school grant) or one or more schools in your district (if you received a district grant) have been chosen for this study. I believe it is important for you to understand the nature of this study and how the data will be used.

What is the nature of the study?

The purpose of the study is to determine the type or nature of the classroom instruction that is now going on in the Gates grantee schools. We already have one measure of that instruction gathered through the teacher questionnaires last year. Those results present one picture of the instruction. We think that it is important to get a second picture through actual classroom observations. Obviously, we cannot observe every class in every school to get a complete picture. But by observing a representative sample of classes in a school we can get a general picture of the type of instruction that goes on in the school. By observing in a representative sample of schools, we can therefore get a general picture of the type of instruction that is going on in all of the Gates grantee schools. We are focusing on this *aggregate* picture, and **not** the results of any one school.

How were the schools selected?

The schools were selected based on several criteria, such as grade levels and results of the teacher questionnaire, to provide a representative sample of schools for meeting the purpose of the research.

Who is conducting the study?

Dr. Carol Stuen Brown is directing the study. She is a former elementary teacher, research instructor, and now works full-time on Gates evaluation activities. She will lead a four-person research team going out to the schools. Some schools will have a two-person team conducting the research, and other schools will have all four team members visiting the school, depending on school size. All of the researchers are former classroom teachers.

What is involved in the classroom observations?

Generally, one researcher will visit a classroom for about 30 minutes. We want to visit classrooms to observe only social studies, language arts, and math and science lessons. In elementary schools we will visit only 3rd grade and above; K-2 will not be involved in this particular part of the evaluation. We want to observe "typical" lessons in those areas, so teachers should not do any special preparation or work for the observations. Therefore, on the part of the teachers there is no extra work or involvement other than having a visitor in the classroom for that short period of time. The observations will be conducted over a one or two day period, depending on the size of the school.

What kind of data will be collected?

Observers will be using a classroom observation instrument that focuses on the general areas of instructional approach used, student work products, and depth of learning. Basically, there will be simple descriptive data in these areas for each classroom observation period. We prefer that teachers **not** know the specifics of the instrument ahead of time because that knowledge may influence or change their behavior or the lessons observed. We will be happy to share this with you after the process is over.

How will the data be reported?

It is important to note this is a broad *research project* that is part of the overall Gates Education Initiatives evaluation, and not specifically a part of the evaluation of your school or district. The results of the research will certainly be instructive for all of the schools and districts, and those results will be made available to all of you. However, individual school data will not be made available to anyone other than that particular school, and then only in descriptive form for the entire school (not individual teachers) and only if the school requests it.

We will be using and reporting the data out primarily in the *aggregate* for all of the schools. While there may be some type of descriptive procedure used, such as a scatter plot of school scores, **no school identification will be possible**.

If there are questions from teachers and principals about how these findings will be reported and used, please assure them of the following:

> **No individual classroom observation results will be reported or available to anyone within the school, within the district, or external to the district.**
>
> **No individual school composite results will be reported or available to anyone within the district or external to the district.**

When will all of this happen?

School visits will begin toward the end of October and continue through mid-January, 2002.

Do we have to participate?

Yes, this is part of the evaluation requirements districts and schools agreed to in the grant contracts. However, if we have chosen a school that has some extraordinary or extenuating circumstance that makes the validity of the observation study questionable in that school, please let me know and we can discuss the situation.

What happens next?

Evaluation coordinators should inform the schools that they have been selected for the study and what it entails. It is quite acceptable that copies of this document be used for that purpose. If there are questions or concerns, please call me at your earliest convenience.

Carol Stuen Brown will begin contacting evaluation coordinators sometime during the second week of October to arrange school visits. She will work with coordinators to schedule day(s) for each of the visits, but will leave it to the coordinator to set up an agenda for each observation day; that is, the specific order of classroom observation times within each school will be the responsibility of the coordinator. It will be to your benefit to prepare a list of 5 "good" days for each of your selected schools before Carol calls. Remember, the observations will only involve language arts, math, social studies and science classes and at the elementary level will be limited to 3rd grade and above.

Appendix D Subject Area Classes Observed by School

Appendix D Subject Area Classes Observed by School

School	LA	Math	Science	Social Studies	Integrated
High Schools					
Port Angeles HS	19	11	14	15	
Havermale Alt-Spokane	4	2	1	0	5
Southridge HS-Kennewick	13	11	12	11	
Mt. View HS-Evergreen	14	13	10	20	
Evergreen HS-Evergreen	9	10	13	9	
Sealth HS-SSD	9	7	7	10	
Garfield HS-SSD	10	11	8	8	
Squalicum HS-BSD	6	8	6	7	
Nova Alternative HS-SSD	3	2	6	1	
High School Totals	**87**	**75**	**77**	**81**	**5**
Technical Schools	**LA**	**Math**	**Science**	**Social Studies**	**Integrated**
Tri-Tech Skills Center-Kennewick	0	0	0	0	15
Clark County Skills Center	0	0	0	0	6
Technical Schools Totals	**0**	**0**	**0**	**0**	**21**
Middle Schools	**LA**	**Math**	**Science**	**Social Studies**	**Integrated**
Park MS-Kennewick	3	6	7	9	
Garry MS-Spokane	10	5	4	2	
Glover MS-Spokane	6	6	5	7	
Wy'East MS-Evergreen	1	6	2	10	
Environmental & Adventure School-LWSD	3	1	4	4	
Aki Kurose MS-SSD	9	4	0	7	
COHO-NOMS	6	3	6	6	
Pathfinder	8	6	2	2	
Middle School Totals	**46**	**37**	**30**	**47**	
Elementary Schools	**LA**	**Math**	**Science**	**Social Studies**	**Integrated**
Southgate Elem-Kennewick	7	11	3	0	
View Ridge-Kennewick	11	6	1	1	
Washington Elem-Kennewick	1	2	7	1	
Linwood Elem-Spokane	6	9	2	0	
Prosser Heights Elem-Prosser	10	5	4	3	
Sunset Elementary-Evergreen	3	5	0	1	9
Orchards Elementary-ESD	8	3	1	3	2
B.F. Day Elementary-SSD	4	5	0	0	1
Lowell Elementary-SSD	1	4	1	1	5
Rainier View Elementary-SSD	6	5	4	0	
Happy Valley Elem-BSD	9	7	1	1	
John Hay Elem-SSD	5	3	0	4	
Sacajawea Elem-SSD	12	3	0	0	

Schmitz Park Elementary	6	6	1	0	1
Madrona Elementary	5	5	1	4	
Elementary Totals	**94**	**79**	**26**	**19**	**18**
Totals by Subject	**227**	**191**	**133**	**147**	**44**

Appendix E Inter-rater Agreement and Reliability Statistics by Observation

Appendix E Inter-rater Agreement and Reliability Statistics by Observation

High Schools	Agree 0-1	2	3	4	Reliability coefficient
Squalicum HS-BSD	89%	11%	0%	0%	.61*
Southridge HS-KSD	81%	15%	4%	0%	.63
Southridge HS-KSD	89%	11%	0%	0%	.80
Mt. View HS-ESD	85%	11%	4%	0%	.84
Mt. View HS-ESD	89%	11%	0%	0%	.35
Evergreen HS-ESD	85%	11%	4%	0%	.61
Garfield HS-SSD	96%	4%	0%	0%	.95
Garfield HS-SSD	96%	4%	0%	0%	.74
Sealth HS-SSD	100%	0%	0%	0%	.87
Sealth HS-SSD	100%	0%	0%	0%	.94
Sealth HS-SSD	100%	0%	0%	0%	.94
PortAngeles HS-PASD	74%	19%	7%	0%	.10
PortAngeles HS-PASD	96%	4%	0%	0%	.86*
PortAngeles HS-PASD	96%	4%	0%	0%	.85
PortAngeles HS-PASD	100%	0%	0%	0%	.80
PortAngeles HS-PASD	96%	4%	0%	0%	.39
Havermale HS-SSD	85%	4%	11%	0%	.74
Mt.View-ESD	63%	22%	15%	0%	.49
Nova Alternative HS-SSD	100%	0%	0%	0%	.98
AVERAGE	**91%**	**7%**	**2%**	**0%**	
			*Indicates an average coefficient for an observation with multiple observers.		

Middle Schools	Agree 0-1	2	3	4	Reliability coefficient
Aki Kurose MS-SSD	100%	0%	0%	0%	.67
Aki Kurose MS-SSD	96%	4%	0%	0%	.89
Park Middle-KSD	100%	0%	0%	0%	.82
Garry MS-SSD	96%	4%	0%	0%	.86
Glover MS-SSD	96%	4%	0%	0%	.91
Wy'East MS-ESD	92%	4%	4%	0%	.61
Envir. & Ad.Sch.-LWSD	93%	7%	0%	0%	.76
Envir. & Ad. Sch.-LWSD	100%	0%	0%	0%	.96
COHO-NOMS-SSD	89%	11%	0%	0%	.85
COHO-NOMS-SSD	96%	4%	0%	0%	.71
COHO-NOMS-SSD	96%	4%	0%	0%	.84
Pathfinder-SSD	100%	0%	0%	0%	.94
Pathfinder-SSD	100%	0%	0%	0%	.93
Pathfinder-SSD	100%	0%	0%	0%	.84
Pathfinder-SSD	100%	0%	0%	0%	.84
Pathfinder-SSD	100%	0%	0%	0%	.94
Pathfinder-SSD	85%	7%	7%	0%	.97
Average	**96%**	**3%**	**1%**	**0%**	

Elementary Schools	**Agree 0-1**	**2**	**3**	**4**	**Reliability coefficient**
Ridgeview-KSD	89%	11%	0%	0%	.72
Southgate-KSD	88%	11%	0%	3%	.80
Washington-KSD	96%	4%	0%	0%	.85
Washington-KSD	85%	4%	11%	0%	.76
Linwood-SSD	89%	11%	0%	0%	.81
Sunset Elementary-ESD	81%	15%	4%	0%	.68
Orchards Elementary-ESD	88%	19%	0%	0%	.71
Lowell Elementary-SSD	74%	19%	7%	0%	.77
B.F. Day Elem-SSD	100%	0%	0%	0%	.81
Rainier View Elem-SSD	100%	0%	0%	0%	.80
Rainier View Elem-SSD	96%	4%	0%	0%	.78
Rainier View Elem-SSD	100%	0%	0%	0%	.95
Rainier View Elem-SSD	96%	4%	0%	0%	.62
Rainier View Elem-SSD	100%	0%	0%	0%	.72
Rainier View Elem-SSD	100%	0%	0%	0%	.84
Rainier View Elem-SSD	100%	0%	0%	0%	.91
John Hay Elem-SSD	100%	0%	0%	0%	.98
John Hay Elem-SSD	100%	0%	0%	0%	.90
John Hay Elem-SSD	100%	0%	0%	0%	.95
John Hay Elem-SSD	100%	0%	0%	0%	.99
John Hay Elem-SSD	100%	0%	0%	0%	.98
Sacajawea Elem-SSD	100%	0%	0%	0%	.79
Sacajawea Elem-SSD	100%	0%	0%	0%	.94
Sacajawea Elem-SSD	93%	7%	0%	0%	.87
Happy Valley Elem-BSD	100%	0%	0%	0%	.91
Happy Valley Elem-BSD	93%	7%	0%	0%	.73
Schmitz Park-SSD	100%	0%	0%	0%	.91
Schmitz Park-SSD	100%	0%	0%	0%	.92
Madrona-SSD	100%	0%	0%	0%	.79
Madrona-SSD	100%	0%	0%	0%	.91
Madrona-SSD	100%	0%	0%	0%	.64
Madrona-SSD	100%	0%	0%	0%	.89
AVERAGE	**96%**	**4%**	**1%**	**0%**	

Appendix F: Sample Constructivist and Non-Constructivist Lessons and Scoring

Holistic Observation Rating: 4
Average of Components 1-6: 2.5
Subject: Science
Grade Level: High School
Topic: Navigation

Background

As the observation began, small groups of students were working together at tables on a navigation project. The teacher was at his desk, answering questions and monitoring student progress. There was a quiet buzz among students, who appeared to be absorbed in their work. Rather than explain the project himself, the teacher asked one of the students sitting nearby to review the objective of the project and the work they had done to date. This student proceeded to give a clear and detailed overview of the activity.

After being given a marine navigational map, students were instructed to design a path to move a tanker from Point A to Point B. In addition to accounting for water currents, water depth, and tides, students were also asked to consider such factors as water density (the waterway connected salt and fresh water sources), ferry schedules, and the mass of the tanker, which left empty and later picked up cargo.

Component 1: Work shows evidence of conceptual understanding, not just recall. (Appropriate methods, fundamental concepts and vocabulary, construction of knowledge, and elaborated conceptual communication)
Score: 4

To complete the project, students used a variety of tools and methods, including calculators, a navigational map, measuring devices. Discussions were substantive and thoughtful, and it was clear that the students took the task seriously. Vocabulary was relevant and appropriate to the task (density, mass, distance, speed, ratio, etc) and the depth of the conversations certainly contributed to their conceptual understanding.

Component 2: Students are engaged in activities to develop understanding and create personal meaning through reflection. (Use of appropriate learning strategies, self-evaluation and revision of work, consideration of alternatives, intentional reflection, focused feedback from the teacher, substantive conversation)
Score: 3.3

Feedback from the teacher as well as student conversations both led to reflection and understanding. Through discussion with group members and/or the teacher, students pondered alternative solutions to their task and revised their work based as they felt it was necessary. The nature of the project was such that students were required to think critically, and indeed there was no way the students could have completed it *without* a certain amount of analysis and higher-level thinking. Feedback to students was impressive in the way it probed and challenged their thinking, and the conversations between students were substantive, serious, and appeared to be "typical" for this class.

Component 3: Apply knowledge in real world contexts. (Knowledge is connected with relevant personal experiences, knowledge is connected across disciplines and/or to real world problems, community resources are involved, student work is produced for an audience beyond the class, students connect with the world outside school via field experiences or technology)
Score: 1.2

While students did not take their project beyond the walls of the classroom, the task did rely on "real world" resources, such as navigation maps, tidal action, and ferry schedules. In addition, the task encouraged an understanding and appreciation for the practical issues related to marine transportation.

Component 4: Students are engaged in active participation, exploration, and research. (Student work collaboratively, generate their own ideas, questions and hypotheses, plan and/or carry out independent research, and independently access and use print media, equipment or technology)
Score: 2.5

Students in this class were involved in a lesson that exemplified collaboration. Their discussions were thoughtful and serious, and their interactions were positive and productive. Although unstated, there appeared to be an expectation that students would generate their own ideas, hypotheses, and questions.

Component 5: Teacher uses diverse experiences of students to build effective learning. (Teacher accesses prior knowledge, accommodates student strengths and needs, and builds the lesson on diverse cultural tradition, student interests or experiences)
Score: 0

There was no activation of prior knowledge during observation, nor were there any obvious accommodation of individual differences. And while the lesson appeared to interest students, it was not built on their diverse experiences.

Component 6: Students are presented with a challenging curriculum designed to develop depth of understanding (Lesson presented emphasizes conceptual understanding, and the central ideas an concepts of the subject are covered in depth)
Score: 4

This lesson, both in terms of planning and facilitation (teacher) and understanding (students), was an excellent example of how a challenging, conceptually rich curriculum can be designed and presented to students. The teacher provided enough background and structure to engage and direct students, while leaving the task "fuzzy" enough so that students had to stretch their thinking. Not only did students appear to be collaborating and learning, but they were also motivated and challenged by the task.

Holistic Observation Rating: 3
Average of Components 1-6: 2.15
Subject: Reading
Grade Level: 5
Topic: Literature/Novel Study

Background

During the time of this observation, three groups of students were working on various reading tasks: two groups were doing assignments related to their novels, and one group was involved in a literature discussion group with the teacher. Each group was reading a different novel, and while not stated directly by the teacher, it was assumed by the observers that novels were selected to accommodate different reading levels. The observation focused primarily on the discussion group. It is useful to note, however, that the two other (independent) groups were involved in meaningful projects related to their books, since this is often a weakness of differentiated group instruction. Some students were working on 5-paragraph essays (responding to a series of questions about their novels), some were illustrating their essays, and others were designing a 20-word crossword using vocabulary words of their choice from the novel. Nearly every student appeared to be engaged and focused on a meaningful task.

As the observation began, a literature circle group of 11 students was discussing their novel. Listening to the discussion, it appeared that all students had read the required chapters, evidenced by the fact that all students contributed ideas and examples.

Component 1: Work shows evidence of conceptual understanding, not just recall. (Appropriate methods, fundamental concepts and vocabulary, construction of knowledge, and elaborated conceptual communication)
Score: 3.4

Both the novel and a discussion guide were used to facilitate an analysis of the story, and several different strategies were employed to develop conceptual understanding. These included the consideration of alternatives, prediction, comparing and contrasting, and a discussion of cause and effect. Vocabulary was relevant and appropriately used (setting, illustration, communicate, character, flow, etc). Elaborated communication between students and teacher was important in how they were able to make sense of the story. The following examples of the questions posed by the teacher illustrate the way in which discussion was generated:

T: "Let's review the sequence of the story."

T: "Who are the characters in this story? How would you describe them…what about physical attributes?"

T: "What do we mean by topography?"

Component 2: Students are engaged in activities to develop understanding and create personal meaning through reflection. (Use of appropriate learning strategies, self-evaluation and revision of work, consideration of alternatives, intentional reflection, focused feedback from the teacher, substantive conversation)
<u>Score: 2.8</u>

Various strategies were used to help students understand the stories. For those working independently, computers, reading guides, art materials, reading logs, graph paper, and books were in evidence. Students working with the teacher, as noted previously, used a discussion guide and their novel. Intentional reflection was an important part of the literature circle discussion, as were the reading guide questions. For example, the teacher asked students to reflect on their own personal trials, a theme that emerged from the reading. Students did this, and some volunteered to share their thoughts with the group. In another instance, they were asked to think about what the author meant by "traps," again a reference to a theme in the story. The entire discussion was a substantive conversation that encouraged critical thinking, helped students make sense of the story, and prompted them to consider some important aspects of the human experience.

Component 3: Apply knowledge in real world contexts. (Knowledge is connected with relevant personal experiences, knowledge is connected across disciplines and/or to real world problems, community resources are involved, student work is produced for an audience beyond the class, students connect with the world outside school via field experiences or technology)
<u>Score: .8</u>

Students were provided with several opportunities to meaningfully reflect on, and then share, their personal experiences as related to the story. Aside from this, the lesson did little to connect the novel to real world experiences.

Component 4: Students are engaged in active participation, exploration, and research. (Student work collaboratively, generate their own ideas, questions and hypotheses, plan and/or carry out independent research, and independently access and use print media, equipment or technology)
<u>Score: 1.2</u>

While a few students worked collaboratively, most were focused on their own, individual projects. They did, however, generate their own ideas and questions, particularly those participating in the literature circle. Students did not engage in independent research during the observation, and computers were used primarily for word-processing.

Component 5: Teacher uses diverse experiences of students to build effective learning. (Teacher accesses prior knowledge, accommodates student strengths and needs, and builds the lesson on diverse cultural tradition, student interests or experiences)
<u>Score: 1.7</u>

This lesson built on student experience and interest, to a certain degree ("Who has read………"), and the learning was differentiated in that different groups read different novels.

Reading guides for all novels encouraged critical thinking and related to student experience. This was a strength of the lesson.

Component 6: Students are presented with a challenging curriculum designed to develop depth of understanding (Lesson presented emphasizes conceptual understanding, and the central ideas an concepts of the subject are covered in depth)
Score: 3

This lesson was rated "3" based on several factors. First, the lesson emphasized conceptual understanding because of the ways in which students were asked to respond to, and analyze their novels. Comprehension questions (both for written response and for discussion) encouraged higher-level thinking and personal reflection. Secondly, the independent work in which students were engaged was clearly motivating, and students were producing high quality products (reading logs, illustrations, crosswords, and essays). Finally, there was a depth to the literature analysis that was impressive. Elements of the story were discussed seriously (theme, setting, character, climax, etc) and students were encouraged to relate the story themes to their own life experience.

Holistic Observation Rating: 2
Average of Components 1-6: 1.8
Subject: Literacy
Grade Level: 3
Topic: Writing

Background

This writing lesson focused on students' experiences in P.E. class (basketball), which were ultimately to be included in a school newsletter. They were shown how to make a "grid" to be used in organizing their thoughts. These thoughts would subsequently be expanded into a brief article. A class discussion, led by the teacher, generated ideas on what types of information would be appropriate for the article, with students offering suggestions.
T: "What is important when you play basketball with your friends?"

S: "Making baskets"
S: "The ball"
S: "Being a good dribbler"

Eventually there was agreement about the basic elements of the game: *Equipment, Skills, Rules,* and *Where to Play*. These became categories for the "grid," which was a piece of paper folded into four sections. (Students were also told that if these categories did not work for them, they could choose different ones. It appeared that all students used the identified categories, however). Students were then told to write only single words or phrases in each section of the grid, enough to remind them of what they wanted to say, but without writing entire sentences. The intent was to have students develop an outline, of sorts, from which they could write an article.

As the lesson unfolded, it appeared that a number of students did not understand the task. Most students were writing complete sentences, which were difficult to fit in the boxes, and more importantly, which were not particularly insightful. For example, in the section about Rules, several students wrote, "There are rules." While the task encouraged students to write from their own experience and perspective, the product turned out to be fairly limited.

Component 1: Work shows evidence of conceptual understanding, not just recall. (Appropriate methods, fundamental concepts and vocabulary, construction of knowledge, and elaborated conceptual communication)
Score: 1.6

Students used the "grid" organizer to prepare for, and guide their writing task, which probably led to some degree of conceptual understanding (of the writing process). There did not appear to be a clear understanding of outlining, summarizing or using key words, however, and as a result, use of relevant vocabulary by the students was minimal. As noted previously, the depth of their thinking was limited to basic and unoriginal reflections about playing basketball, and the focus on using key words was not generally understood.

Component 2: Students are engaged in activities to develop understanding and create personal meaning through reflection. (Use of appropriate learning strategies, self-evaluation and revision of work, consideration of alternatives, intentional reflection, focused feedback from the teacher, substantive conversation)
Score: 1.8

The grid appeared to be an appropriate organizer for this lesson, although students needed more guidance on how to use it. The impact was therefore not as powerful as it might have been. Students did revise their work, based on feedback from the teacher and a parent helper, who attempted to "re-teach" the outlining strategy with individual students. There was also some attempt to have students reflect on their own learning, specifically related to their experiences playing basketball in P.E. ("Think about what is important when you play basketball………"). The teacher and parent helper did provide feedback to students; however the feedback did not probe or extend their conceptual understanding, but rather answered specific questions. These exchanges could not be considered "substantive conversation."

Component 3: Apply knowledge in real world contexts. (Knowledge is connected with relevant personal experiences, knowledge is connected across disciplines and/or to real world problems, community resources are involved, student work is produced for an audience beyond the class, students connect with the world outside school via field experiences or technology)

Score: 2.2

The lesson related directly to student experience (P.E. class and basketball) and also connected writing to a real-world event (school newsletter). However, as the teacher pointed out, not all 24 of the articles could be included ("Will there be enough space in the newsletter for *everyone's* article?" "Nooooooo…."). The task also involved a community resource in the form of a parent volunteer.

Component 4: Students are engaged in active participation, exploration, and research. (Student work collaboratively, generate their own ideas, questions and hypotheses, plan and/or carry out independent research, and independently access and use print media, equipment or technology)
Score: 1

While students were allowed and even encouraged to generate their own ideas, there was little collaboration, no independent research, and no independent use of media.

Component 5: Teacher uses diverse experiences of students to build effective learning. (Teacher accesses prior knowledge, accommodates student strengths and needs, and builds the lesson on diverse cultural tradition, student interests or experiences)
Score: 2.3

The teacher activated prior knowledge by asking students to recall previous writing assignments done for the school newsletter, and to recall their experiences in P.E. class. The students found this interesting and motivating, and thus at the outset approached the lesson with enthusiasm. There was also an attempt to differentiate instruction when the teacher allowed students to create categories other than those suggested. The lesson was definitely built on student interest and experience.

Component 6: Students are presented with a challenging curriculum designed to develop depth of understanding (Lesson presented emphasizes conceptual understanding, and the central ideas an concepts of the subject are covered in depth)
Score: 2

The degree to which students had the opportunity to develop conceptual understanding was limited. They did use an organizer for their writing, but generally it was not used effectively nor did it lead to high quality writing.

Holistic Observation Rating: 1
Average of Components 1-6: .7
Subject: Science
Grade Level: High School
Topic: Water Cycle

Background

At the outset of the observation, the classroom lights were off, students were seated at lab tables, and the teacher was in the front of the room discussing diagrams shown from the overhead projector. Students appeared to be taking notes. It became clear after several minutes of watching and listening that the diagrams being discussed were drawings that had been done by students describing (in pictures and words) various elements of the water cycle. It also appeared that students had previously read and taken notes on water cycle material from their textbooks. When reminded by the teacher to take notes, one student stated that they "already had these notes." The teacher suggested that it would be beneficial to take notes again since there was a quiz coming up shortly. About a third of the class was taking notes while the rest of the students were talking to each other, writing notes to each other, or working on other assignments. As each diagram was presented, the teacher again reminded students to take notes, and then asked the "author" to briefly review their work.

At the end of the lesson, the teacher did a brief review and then passed out a quiz that consisted of questions on the material discussed in class. Students were required to draw diagrams and write brief explanations. They were encouraged to use their notes.

Component 1: Work shows evidence of conceptual understanding, not just recall. (Appropriate methods, fundamental concepts and vocabulary, construction of knowledge, and elaborated conceptual communication)
Score: 1.2

Methods used in this lesson included teacher lecture from overhead transparencies and student note-taking. Transparencies were student diagrams and descriptions of elements of the water cycle. In most cases when relevant vocabulary was used during the lesson, it was used by the teacher. When students *did* use relevant vocabulary, it was because the teacher prompted them. Students had no opportunities to manipulate information or construct knowledge, nor did any elaborated communication take place. Exchanges between teacher and students were limited to knowledge and comprehension level questions and answers, which did little to probe or extend student understanding. In fact, there was relatively little engagement on the part of students at all.

Component 2: Students are engaged in activities to develop understanding and create personal meaning through reflection. (Use of appropriate learning strategies, self-evaluation and revision of work, consideration of alternatives, intentional reflection, focused feedback from the teacher, substantive conversation)

Score: .3

Simply stated, there were few opportunities for students to create personal meaning. Neither was there any challenge to reflect on the material. And while note taking *can* be an effective learning strategy, in this particular lesson it was a repeat of work already done. Students were not encouraged to think at a higher level, nor were any probing questions asked. Two examples follow which illustrate the level of questioning: throughout the lesson:

Teacher, pointing to a diagram of a ground spring: "Where does the water for this spring come from?"
Student: "Underground."
T: "Yes….remember the water table………what does that have to do with the spring?"
S: "It fills the spring."
T: "What happens when the water table is down here?"
S: "There's no water for the spring."

Teacher, showing a diagram of underground layers: "What is permeable?"
Student #1: "I…..I don't know."
Student #2: "I'm not sure……wait, I know. It's permanent rock."
Student #3: "Permeable is where water can flow through."
T: " And what is impermeable?"
Student #3: "That's where it can't flow through."

Component 3: Apply knowledge in real world contexts. (Knowledge is connected with relevant personal experiences, knowledge is connected across disciplines and/or to real world problems, community resources are involved, student work is produced for an audience beyond the class, students connect with the world outside school via field experiences or technology)
Score: 1.0

Three times during the lesson there were references to real world and/or relevant experiences, twice by the teacher and once by a student. For example, a student brought up the problem of contaminated water in foreign countries and the consequences for travelers. The teacher used a bank account as an analogy for the water cycle, where the input (groundwater: monetary deposits) must equal the output (precipitation: spending). The teacher also discussed the flooding situation of a local river and the consequences for residents.

Component 4: Students are engaged in active participation, exploration, and research. (Student work collaboratively, generate their own ideas, questions and hypotheses, plan and/or carry out independent research, and independently access and use print media, equipment or technology)
Score: 0

Students were relatively uninvolved in the lesson, and there was no collaboration. Some took notes, but there was no evidence that any students generated ideas, hypotheses or questions. Technology was not used, nor were students involved in any independent research.

Component 5: Teacher uses diverse experiences of students to build effective learning. (Teacher accesses prior knowledge, accommodates student strengths and needs, and builds the lesson on diverse cultural tradition, student interests or experiences)
Score: .7

Aside from the examples previously mentioned where the teacher and one student connected the lesson to the real world, there were no attempts to accommodate individual differences or diverse experiences.

Component 6: Students are presented with a challenging curriculum designed to develop depth of understanding (Lesson presented emphasizes conceptual understanding, and the central ideas an concepts of the subject are covered in depth)
Score: 1.0

The lesson can be summed up as a low level review of the elements of the water cycle. While the basic elements of the water cycle were indeed covered, the information was not unlike a lesson one might see in a 5th or 6th grade classroom. Conceptual understanding did not appear to be an expectation, and the material was covered at only the most superficial level. Reasoning, reflection, probing questions, and self-evaluation were absent. When students were unsure of a correct answer, the teacher prompted them until they got it right. Even then, however, it seemed not to be effective. When the teacher posed one final review question before handing out the quiz, the first two students called on did not know the answer.

Holistic Observation Rating: 1
Average of Components 1-6: .3
Subject: Science
Grade Level: High School
Topic: Unity and Diversity

Background

When the observers arrived in the classroom, the teacher was working on the computer in an office separated from the rest of the class. After making introductions, the teacher said to "go ahead and walk around, sure." The classroom was a tiered science lab, with students seated at lab tables on four different levels. At the front of the room there was a white board and a lab bench. When asked what they were working on, a couple students volunteered that they answering and discussing questions from the textbook, and writing responses in their journals. The chapter focused on identification of plants (unique characteristics, similarities, and how scientists identify various plant species). Two groups of students (5-6 students) did appear to be working on the assignment. The rest of the class had their books open, but were talking about their plans for the weekend, upcoming dates, dances, shopping experiences and the like. The teacher did not come out of the office during the observation.

Component 1: Work shows evidence of conceptual understanding, not just recall. (Appropriate methods, fundamental concepts and vocabulary, construction of knowledge, and elaborated conceptual communication)
Score: .2

Some students were using the textbook to answer questions at the end of the chapter. Use of relevant vocabulary was minimal, and there was no evidence of construction of knowledge, solving complex problems or elaborated communication during the observation.

Component 2: Students are engaged in activities to develop understanding and create personal meaning through reflection. (Use of appropriate learning strategies, self-evaluation and revision of work, consideration of alternatives, intentional reflection, focused feedback from the teacher, substantive conversation)
Score: .3

Students did use the book and a journal to record responses, but otherwise there was no attempt on the part of either the teacher or the students to create personal meaning.

Component 3: Apply knowledge in real world contexts. (Knowledge is connected with relevant personal experiences, knowledge is connected across disciplines and/or to real world problems, community resources are involved, student work is produced for an audience beyond the class, students connect with the world outside school via field experiences or technology)
Score: 0

At no time did students appear to make connections between their reading and the real world.

Component 4: Students are engaged in active participation, exploration, and research. (Student work collaboratively, generate their own ideas, questions and hypotheses, plan and/or carry out independent research, and independently access and use print media, equipment or technology)
Score: .5

Although students were talking, in most cases the conversations were not related to school. When students *did* discuss the questions, it was to get an answer, not to share knowledge or critique their work. No independent research occurred.

Component 5: Teacher uses diverse experiences of students to build effective learning. (Teacher accesses prior knowledge, accommodates student strengths and needs, and builds the lesson on diverse cultural tradition, student interests or experiences)
Score: 0

This did not happen.

Component 6: Students are presented with a challenging curriculum designed to develop depth of understanding (Lesson presented emphasizes conceptual understanding, and the central ideas an concepts of the subject are covered in depth)
<u>Score: 1</u>

Although the students had an assignment, it was strictly a matter of finding answers to end-of-chapter questions. The teacher was in another room for the entire 30 minutes of the observation. The only reason for a score of "1" on this component was that the questions in the book were rigorous, and had they been addressed seriously by the teacher and students they could have been the basis for a challenging lesson. As it was, the lesson offered little in the way of conceptual understanding.

Appendix G: Descriptive Statistics and MANOVA Results for Elementary, Middle/Junior, and High Schools

Appendix G Descriptive Statistics and MANOVA Results for Elementary, Middle/Junior, and High Schools

The data from the 669 classroom observations were analyzed using SPSS General Linear Model (Multivariate) with school status as the fixed factor and with the first 6 components and the holistic score on the TAOP as the dependent variables. Component 7 dealing with summative assessment was excluded because of the limited number of cases with scores on that variable. The Wilks' *Lambda* and all of the between-subjects univariate tests were significant. Tamhane or LSD p*ost hoc* results were used depending on the Levene's Test of Equality of Error Variances outcome. Of the 21 comparisons made among the elementary, middle/junior, and high school groups, none of the comparisons were significant at the .01 level, and only 1 comparison was significant at the .05 level. None of the 7 comparisons made between the alternative school and voctech schools were significant at the .05 level. Of the 42 comparisons made among the alternative/voctech schools and the elementary/middle-junior/high schools, 38 significant differences were found, 29 at the .01 level and 9 at the .05 level. In all of the comparisons with significant differences, the alternative/voctech schools' scores were higher than were the scores of the elementary/middle-junior/high schools.

Descriptive Statistics

	level	Mean	Std. Deviation	N
SECT1	**elementary**	2.0640	1.04924	200
	middle/jr.high	2.0125	1.18574	136
	high school	2.1405	1.12825	268
	alternative	2.8636	.94493	22
	voctech	3.1238	.83780	21
	Total	2.1465	1.12441	647
SECT2	**elementary**	1.6146	1.02002	200
	middle/jr.high	1.5282	1.09280	136
	high school	1.5605	1.05434	268
	alternative	2.4735	1.12258	22
	voctech	2.5556	.89339	21
	Total	1.6338	1.07309	647
SECT3	**elementary**	.5625	.68041	200
	middle/jr.high	.7096	.91359	136
	high school	.6468	.82437	268
	alternative	1.7500	1.04733	22
	voctech	2.6571	1.05099	21
	Total	.7367	.91408	647
SECT4	**elementary**	.8569	.76916	200
	middle/jr.high	1.0018	1.10376	136
	high school	.8741	.92723	268
	alternative	1.7386	1.19890	22
	voctech	1.9048	1.18183	21
	Total	.9585	.96878	647
SECT5	**elementary**	1.0683	1.01952	200
	middle/jr.high	1.0208	.98746	136
	high school	.8083	.88066	268
	alternative	2.4318	1.35172	22
	voctech	1.7619	1.35459	21
	Total	1.0195	1.03238	647
SECT6	**elementary**	2.1325	1.28333	200

	middle/jr.high	2.0110	1.25789	136
	high school	2.2671	1.21931	268
	alternative	2.8523	.88862	22
	voctech	2.9286	1.20712	21
	Total	2.2130	1.25125	647
Overall conclusion: How constructivist was this lesson?	**elementary**	2.4225	1.01136	200
	middle/jr.high	2.2206	1.05719	136
	high school	2.3476	1.00927	268
	alternative	3.0909	1.01929	22
	voctech	3.2381	.76842	21
	Total	2.3982	1.03387	647

Box's Test of Equality of Covariance Matrices[a]

Box's M	235.143
F	1.942
df1	112
df2	18461.168
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

[a] Design: Intercept+VAR00001

Multivariate Tests(d)

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power(a)
Intercept	**Pillai's Trace**	.741	259.677(b)	7.000	636.000	.000	.741	1817.738	1.000
	Wilks' Lambda	.259	259.677(b)	7.000	636.000	.000	.741	1817.738	1.000
	Hotelling's Trace	2.858	259.677(b)	7.000	636.000	.000	.741	1817.738	1.000
	Roy's Largest Root	2.858	259.677(b)	7.000	636.000	.000	.741	1817.738	1.000
VAR00001	**Pillai's Trace**	.345	8.628	28.000	2556.000	.000	.086	241.570	1.000
	Wilks' Lambda	.684	9.087	28.000	2294.553	.000	.090	228.133	1.000
	Hotelling's Trace	.419	9.485	28.000	2538.000	.000	.095	265.574	1.000
	Roy's Largest Root	.288	26.300(c)	7.000	639.000	.000	.224	184.097	1.000

a Computed using alpha = .05

b Exact statistic

c The statistic is an upper bound on F that yields a lower bound on the significance level.

d Design: Intercept+VAR00001

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
SECT1	3.257	4	642	.012
SECT2	1.365	4	642	.245
SECT3	4.833	4	642	.001
SECT4	6.187	4	642	.000
SECT5	8.008	4	642	.000
SECT6	1.730	4	642	.142
Overall conclusion: How constructivist was this lesson?	2.486	4	642	.042

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+VAR00001

Tests of Between-Subjects Effects									
Source	**Dependent Variable**	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**	**Noncent. Parameter**	**Observed Power(a)**
Corrected Model	**SECT1**	35.185(b)	4	8.796	7.226	.000	.043	28.902	.996
	SECT2	36.383(c)	4	9.096	8.254	.000	.049	33.015	.999
	SECT3	108.376(d)	4	27.094	40.322	.000	.201	161.290	1.000
	SECT4	36.425(e)	4	9.106	10.259	.000	.060	41.035	1.000
	SECT5	67.894(f)	4	16.973	17.558	.000	.099	70.232	1.000
	SECT6	27.372(g)	4	6.843	4.464	.001	.027	17.858	.940
	Holistic	30.464(h)	4	7.616	7.408	.000	.044	29.632	.997
Intercept	**SECT1**	1364.534	1	1364.534	1120.881	.000	.636	1120.881	1.000
	SECT2	867.721	1	867.721	787.388	.000	.551	787.388	1.000
	SECT3	366.605	1	366.605	545.597	.000	.459	545.597	1.000
	SECT4	372.448	1	372.448	419.587	.000	.395	419.587	1.000
	SECT5	460.656	1	460.656	476.523	.000	.426	476.523	1.000
	SECT6	1361.625	1	1361.625	888.351	.000	.580	888.351	1.000
	Holistic	1625.308	1	1625.308	1580.906	.000	.711	1580.906	1.000
VAR00001	**SECT1**	35.185	4	8.796	7.226	.000	.043	28.902	.996
	SECT2	36.383	4	9.096	8.254	.000	.049	33.015	.999
	SECT3	108.376	4	27.094	40.322	.000	.201	161.290	1.000
	SECT4	36.425	4	9.106	10.259	.000	.060	41.035	1.000
	SECT5	67.894	4	16.973	17.558	.000	.099	70.232	1.000
	SECT6	27.372	4	6.843	4.464	.001	.027	17.858	.940
	Holistic	30.464	4	7.616	7.408	.000	.044	29.632	.997
Error	**SECT1**	781.556	642	1.217					
	SECT2	707.500	642	1.102					
	SECT3	431.382	642	.672					
	SECT4	569.874	642	.888					
	SECT5	620.623	642	.967					
	SECT6	984.029	642	1.533					
	Holistic	660.032	642	1.028					
Total	**SECT1**	3797.688	647						
	SECT2	2470.879	647						
	SECT3	890.886	647						
	SECT4	1200.665	647						
	SECT5	1360.985	647						
	SECT6	4180.097	647						
	Holistic	4411.778	647						
Corrected Total	**SECT1**	816.741	646						
	SECT2	743.883	646						
	SECT3	539.758	646						
	SECT4	606.299	646						
	SECT5	688.517	646						
	SECT6	1011.401	646						
	Holistic	690.496	646						

a Computed using alpha = .05
b R Squared = .043 (Adjusted R Squared = .037)
c R Squared = .049 (Adjusted R Squared = .043)
d R Squared = .201 (Adjusted R Squared = .196)
e R Squared = .060 (Adjusted R Squared = .054)
f R Squared = .099 (Adjusted R Squared = .093)
g R Squared = .027 (Adjusted R Squared = .021)
h R Squared = .044 (Adjusted R Squared = .038)

level					
		Mean	Std. Error	95% Confidence Interval	
Dependent Variable	**level**			**Lower Bound**	**Upper Bound**
SECT1	**elementary**	2.064	.078	1.911	2.217
	middle/jr.high	2.013	.095	1.827	2.198
	high school	2.141	.067	2.008	2.273
	alternative	2.864	.235	2.402	3.326
	voctech	3.124	.241	2.651	3.597
SECT2	**elementary**	1.615	.074	1.469	1.760
	middle/jr.high	1.528	.090	1.351	1.705
	high school	1.561	.064	1.435	1.686
	alternative	2.473	.224	2.034	2.913
	voctech	2.556	.229	2.106	3.005
SECT3	**elementary**	.562	.058	.449	.676
	middle/jr.high	.710	.070	.572	.848
	high school	.647	.050	.548	.745
	alternative	1.750	.175	1.407	2.093
	voctech	2.657	.179	2.306	3.008
SECT4	**elementary**	.857	.067	.726	.988
	middle/jr.high	1.002	.081	.843	1.160
	high school	.874	.058	.761	.987
	alternative	1.739	.201	1.344	2.133
	voctech	1.905	.206	1.501	2.308
SECT5	**elementary**	1.068	.070	.932	1.205
	middle/jr.high	1.021	.084	.855	1.186
	high school	.808	.060	.690	.926
	alternative	2.432	.210	2.020	2.843
	voctech	1.762	.215	1.341	2.183
SECT6	**elementary**	2.133	.088	1.961	2.304
	middle/jr.high	2.011	.106	1.803	2.219
	high school	2.267	.076	2.119	2.416
	alternative	2.852	.264	2.334	3.371
	voctech	2.929	.270	2.398	3.459
Overall conclusion: How constructivist was this lesson?	**elementary**	2.422	.072	2.282	2.563
	middle/jr.high	2.221	.087	2.050	2.391
	high school	2.348	.062	2.226	2.469
	alternative	3.091	.216	2.666	3.515
	voctech	3.238	.221	2.804	3.673

Appendix H: Descriptive Statistics and MANOVA Results for Course Subjects

Appendix H Descriptive Statistics and MANOVA Results for Course Subjects

The data from the 669 classroom observations were analyzed using SPSS General Linear Model (Multivariate) with course subject matter as the fixed factor and with the first 6 components and the holistic score on the TAOP as the dependent variables. Component 7 dealing with summative assessment was excluded because of the limited number of cases with scores on that variable. The Wilks' *Lambda* and all of the between-subjects univariate tests were significant. Tamhane or LSD p*ost hoc* results were used depending on the Levene's Test of Equality of Error Variances outcome. Of the 28 comparisons made between the integrated subject classes and the English, math, science and social studies, 24 significant differences were found, 17 at the .01 level and 7 at the .05 level. In all comparisons the integrated subject classes had higher scores on the TAOP. In the remaining groups of 21 comparisons made between a subject matter class and the other three types of non-integrated subject matter classes, the sciences classes had significantly higher scores on 7 comparisons, the English classes on 5 comparisons, the social studies classes on 1 comparison and the math classes on 0 comparisons.

Descriptive Statistics

	Subject	Mean	Std. Deviation	N
SECT1	**English**	2.1417	1.10055	203
	Math	1.9283	1.06429	166
	Science'	2.3575	1.15154	113
	Social Studies	2.0520	1.16589	125
	Integrated or Other	2.7750	.99040	40
	Total	2.1465	1.12441	647
SECT2	**English**	1.6872	1.06609	203
	Math	1.5224	1.00424	166
	Science'	1.6121	1.11463	113
	Social Studies	1.4893	1.09829	125
	Integrated or Other	2.3375	.93464	40
	Total	1.6338	1.07309	647
SECT3	**English**	.8161	.87802	203
	Math	.2450	.49168	166
	Science'	.7159	.89458	113
	Social Studies	.8896	.87186	125
	Integrated or Other	1.9550	1.20362	40
	Total	.7367	.91408	647
SECT4	**English**	.9672	.92802	203
	Math	.6795	.76829	166
	Science'	1.0863	.99231	113
	Social Studies	.9960	1.09652	125
	Integrated or Other	1.5937	1.06772	40
	Total	.9585	.96878	647
SECT5	**English**	1.2928	1.11173	203
	Math	.7319	.78983	166
	Science'	.8083	.95463	113
	Social Studies	.9080	.95777	125
	Integrated or Other	1.7708	1.26124	40
	Total	1.0195	1.03238	647
SECT6	**English**	2.1363	1.29189	203
	Math	2.0884	1.20746	166

	Science'	2.4358	1.23282	113
	Social Studies	2.1640	1.25125	125
	Integrated or Other	2.6438	1.15732	40
	Total	2.2130	1.25125	647
Overall conclusion: How constructivist was this less?	**English**	2.4269	1.03074	203
	Math	2.1657	.99907	166
	Science'	2.5619	1.05249	113
	Social Studies	2.3240	1.02834	125
	Integrated or Other	2.9875	.86593	40
	Total	2.3982	1.03387	647

Box's Test of Equality of Covariance Matrices(a)

Box's M	316.218
F	2.730
df1	112
df2	124926.109
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a Design: Intercept+SUBJECT

Multivariate Tests(d)

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power(a)
Intercept	**Pillai's Trace**	.840	478.629(b)	7.000	636.000	.000	.840	3350.405	1.000
	Wilks' Lambda	.160	478.629(b)	7.000	636.000	.000	.840	3350.405	1.000
	Hotelling's Trace	5.268	478.629(b)	7.000	636.000	.000	.840	3350.405	1.000
	Roy's Largest Root	5.268	478.629(b)	7.000	636.000	.000	.840	3350.405	1.000
SUBJECT	**Pillai's Trace**	.344	8.595	28.000	2556.000	.000	.086	240.651	1.000
	Wilks' Lambda	.686	9.023	28.000	2294.553	.000	.090	226.529	1.000
	Hotelling's Trace	.414	9.387	28.000	2538.000	.000	.094	262.837	1.000
	Roy's Largest Root	.279	25.461(c)	7.000	639.000	.000	.218	178.229	1.000

a Computed using alpha = .05

b Exact statistic

c The statistic is an upper bound on F that yields a lower bound on the significance level.

d Design: Intercept+SUBJECT

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
SECT1	2.147	4	642	.074
SECT2	1.419	4	642	.226
SECT3	23.744	4	642	.000
SECT4	3.911	4	642	.004
SECT5	12.285	4	642	.000
SECT6	.910	4	642	.458
Overall conclusion: How constructivist was this less?	4.670	4	642	.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+SUBJECT

Tests of Between-Subjects Effects									
Source	**Dependent Variable**	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**	**Noncent. Parameter**	**Observed Power(a)**
Corrected Model	**SECT1**	29.856(b)	4	7.464	6.090	.000	.037	24.359	.987
	SECT2	25.108(c)	4	6.277	5.606	.000	.034	22.426	.979
	SECT3	103.758(d)	4	25.939	38.195	.000	.192	152.781	1.000
	SECT4	31.102(e)	4	7.776	8.679	.000	.051	34.715	.999
	SECT5	58.070(f)	4	14.518	14.784	.000	.084	59.134	1.000
	SECT6	17.107(g)	4	4.277	2.761	.027	.017	11.046	.761
	Overall conclusion: How constructivist was this less?	26.753(h)	4	6.688	6.469	.000	.039	25.876	.991
Intercept	**SECT1**	2398.964	1	2398.964	1957.256	.000	.753	1957.256	1.000
	SECT2	1416.622	1	1416.622	1265.307	.000	.663	1265.307	1.000
	SECT3	404.532	1	404.532	595.663	.000	.481	595.663	1.000
	SECT4	536.569	1	536.569	598.887	.000	.483	598.887	1.000
	SECT5	575.391	1	575.391	585.936	.000	.477	585.936	1.000
	SECT6	2490.927	1	2490.927	1608.353	.000	.715	1608.353	1.000
	Overall conclusion: How constructivist was this less?	2943.235	1	2943.235	2846.821	.000	.816	2846.821	1.000
SUBJECT	**SECT1**	29.856	4	7.464	6.090	.000	.037	24.359	.987
	SECT2	25.108	4	6.277	5.606	.000	.034	22.426	.979
	SECT3	103.758	4	25.939	38.195	.000	.192	152.781	1.000
	SECT4	31.102	4	7.776	8.679	.000	.051	34.715	.999
	SECT5	58.070	4	14.518	14.784	.000	.084	59.134	1.000
	SECT6	17.107	4	4.277	2.761	.027	.017	11.046	.761
	Overall conclusion: How constructivist was this less?	26.753	4	6.688	6.469	.000	.039	25.876	.991
Error	**SECT1**	786.885	642	1.226					
	SECT2	718.775	642	1.120					
	SECT3	436.001	642	.679					
	SECT4	575.196	642	.896					
	SECT5	630.446	642	.982					
	SECT6	994.294	642	1.549					
	Overall conclusion: How constructivist was this less?	663.743	642	1.034					
Total	**SECT1**	3797.688	647						
	SECT2	2470.879	647						
	SECT3	890.886	647						
	SECT4	1200.665	647						
	SECT5	1360.985	647						
	SECT6	4180.097	647						
	Overall conclusion: How constructivist was this less?	4411.778	647						
Corrected Total	**SECT1**	816.741	646						
	SECT2	743.883	646						
	SECT3	539.758	646						
	SECT4	606.299	646						
	SECT5	688.517	646						
	SECT6	1011.401	646						
	Overall conclusion: How constructivist was this less?	690.496	646						

a Computed using alpha = .05
b R Squared = .037 (Adjusted R Squared = .031)
c R Squared = .034 (Adjusted R Squared = .028)
d R Squared = .192 (Adjusted R Squared = .187)
e R Squared = .051 (Adjusted R Squared = .045)
f R Squared = .084 (Adjusted R Squared = .079)
g R Squared = .017 (Adjusted R Squared = .011)
h R Squared = .039 (Adjusted R Squared = .033)

1. Subject

Dependent Variable	Subject	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
SECT1	**English**	2.142	.078	1.989	2.294
	Math	1.928	.086	1.760	2.097
	Science'	2.358	.104	2.153	2.562
	Social Studies	2.052	.099	1.858	2.246
	Integrated or Other	2.775	.175	2.431	3.119
SECT2	**English**	1.687	.074	1.541	1.833
	Math	1.522	.082	1.361	1.684
	Science'	1.612	.100	1.417	1.808
	Social Studies	1.489	.095	1.303	1.675
	Integrated or Other	2.338	.167	2.009	2.666
SECT3	**English**	.816	.058	.703	.930
	Math	.245	.064	.119	.371
	Science'	.716	.078	.564	.868
	Social Studies	.890	.074	.745	1.034
	Integrated or Other	1.955	.130	1.699	2.211
SECT4	**English**	.967	.066	.837	1.098
	Math	.679	.073	.535	.824
	Science'	1.086	.089	.911	1.261
	Social Studies	.996	.085	.830	1.162
	Integrated or Other	1.594	.150	1.300	1.888
SECT5	**English**	1.293	.070	1.156	1.429
	Math	.732	.077	.581	.883
	Science'	.808	.093	.625	.991
	Social Studies	.908	.089	.734	1.082
	Integrated or Other	1.771	.157	1.463	2.079
SECT6	**English**	2.136	.087	1.965	2.308
	Math	2.088	.097	1.899	2.278
	Science'	2.436	.117	2.206	2.666
	Social Studies	2.164	.111	1.945	2.383
	Integrated or Other	2.644	.197	2.257	3.030
Overall conclusion: How constructivist was this less?	**English**	2.427	.071	2.287	2.567
	Math	2.166	.079	2.011	2.321
	Science'	2.562	.096	2.374	2.750
	Social Studies	2.324	.091	2.145	2.503
	Integrated or Other	2.987	.161	2.672	3.303