Jeff C. Marshall, Julie Smart and Robert M. Horton

Tracking Perceived and Observed Growth of Inquiry Practice: A Formative Plan to Improve Professional Development Experiences

Abstract

The authors worked with 22 middle school math and science teachers for one year with the goal of improving the quantity and quality of inquiry-based instruction implemented in the classroom. The professional development experience was framed by the 4E x 2 Instruction Model, which combines key components of inquiry instruction (Engage, Explore, Explain, Extend) with formative assessment and reflective practice integrated into each of the inquiry components. Using the Electronic Quality of Inquiry Protocol (EQUIP), we conducted 102 classroom observations, assessing the teachers on 19 indicators associated with inquiry over 4 constructs: Instruction, Curriculum, Discourse, and Assessment. Combining these results with the teachers' own reports of areas of greatest growth and greatest challenge, we found that teachers improved the most on factors relating to specific instructional techniques. We also found that teachers struggled the most with factors associated with assessment; not only was their growth less in this area, but they fell significantly short of our target of Proficient. As we work to improve instructional practice, we must pay attention not only to the delivery and guidance of material, but also to how assessment can provide the tools that teachers need to modify and enhance their instruction in significant ways.

Key words: Assessment; Growth; Inquiry; Instructional Practice; Instructional Protocol; Mathematics Education; Middle School; Professional Development; Science Education

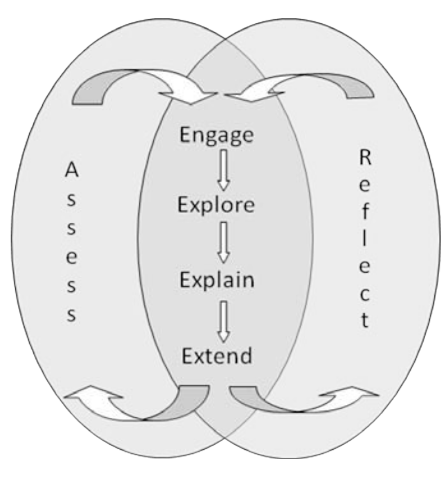
Introduction

In science and mathematics education, a strong effort has been made to offer guidance to practitioners in how to align their teaching practices with inquiry-based reform initiatives (Bybee et al., 2006; Llewellyn, 2007; National Council of Teachers of Mathematics, 1991; National Research Council, 1996, 2000). Many models promote the type of inquiry-based instruction championed by these initiatives, including the Learning Cycle (Karplus, 1977; Renner, Abraham, & Birnie, 1988), the 5E Model (Bybee, 2002), and the 4E x 2 Instructional Model (Marshall, Horton, & Smart, 2009). All three models embrace the constructivist notion that learning begins by understanding and engaging students' prior knowledge. Then, exploration is needed to allow all students to observe, study, and interact with a phenomenon or concept. Finally, meaning is generated as students assimilate these new findings, discoveries, and learnings into existing schema—a Piagetian notion (Piaget, 1970).

The 4E x 2 Model advances prior models by utilizing three critical learning constructs: inquiry-based instruction (National Research Council, NRC, 2000), formative assessment (Black & Wiliam, 1998), and reflective practice (National Board for Professional Teaching Standards, 2006), all of which have been shown to improve student learning (Marshall, Horton, & Smart, 2009). Just as the 5E Model does, the 4E x 2 Instructional Model contains key components of inquiry instruction: Engage, Explore, Explain and Extend. Though assessment is also a key component of both models, in the 4E x 2, Formative Assessment is explicitly distributed throughout each phase of the Model. This is assumed in the 5E Model, but is not clear to all who implement the model. Also, Reflective Practice, which is not explicit in the 5E, is incorporated into each instructional phase of the 4E x 2. See Figure 1.

The 4E x 2 Instructional Model provided the structure for a sustained Professional Development (PD) experience designed to improve the quantity and quality of inquiry instruction

Figure 1. 4E x 2 Instructional Model Visual



a group of teachers led in their classrooms. During this year-long PD experience, math and science teachers from two middle schools first experienced inquiry and then learned to design and implement lessons framed by the 4E x 2 Instructional Model. Our goal was to improve the quality of inquiry instruction that they led in their classrooms.

As the academic year progressed, teachers' success in facilitating inquiry-based instruction was measured by observations made by PD facilitators using the Electronic Quality of Inquiry Protocol (EQUIP) (Marshall, Horton, & White, 2009). These same teachers also reflected on their personal growth and challenges relative to the 19 indicators found on EQUIP.

In this study, we examine how teachers' perceptions of their growth and challenges experienced in implementing inquiry-based practices align with the assessments made by the PD facilitators. In addition, this study examines how components of the current PD model effectively support teacher growth in specific areas and how this model should be adjusted to address areas of teacher challenge. Both teachers and observers used EQUIP to measure the growth of inquiry practice. These measures provided a formative guide for progressive adjustments made to the PD program.

EQUIP contains 19 indicators that comprise four major constructs: Instruction, Discourse, Assessment, and Curriculum. Two of these constructs, Instruction and Assessment, along with the relevant indicators, were central to this study. This decision was reached because teachers perceived Instruction as the area where they experienced the most growth. Conversely, they perceived Assessment as the area where the greatest challenges still remained in implementing inquiry-based instruction. The indicators that comprise these two EQUIP composites are detailed below. Note that all four constructs that frame EQUIP have been discussed and validated in prior work (Marshall, 2009; Marshall, Smart, & Horton, 2010).

Instruction construct.

The five indicators (with the references that validate their inclusion) that comprise the Instruction construct on EQUIP include: (1) *Instructional Strategies* (Abell & Lederman, 2007; Bransford, Brown, & Cocking, 2000; Chiappetta & Koballa, 2006; NRC, 2000), (2) *Order of Instruction* (Abell & Lederman, 2007; Biggs, 1996; Bybee, et al., 2006), (3) *Teacher Role* (Lampert, 1990; Mortimer & Scott, 2003; NRC, 1996; van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001), (4) *Student Role* (Cobb, Wood, & Yackel, 1990), and (5) *Knowledge Acquisition* (Chinn & Brewer, 1998; Mortimer & Scott, 2003). The descriptive rubric used to measure all five Instructional indicators is provided in Appendix A.

Assessment construct.

The second construct explored in this study, Assessment, is framed by five indicators that are pertinent to the facilitation of inquiry-based instruction and include the following: (1) *Prior Knowledge* (Bransford, et al., 2000; Chambers & Andre, 1997), (2) *Conceptual Development* (Driver, Squires, Rushworth, & Wood-Robinson, 1994), (3) *Student Reflection* (Mezirow, 1990; White & Frederiksen, 1998, 2005; Wiggins & McTighe, 1998), (4) *Assessment Type(s)* (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black & Wiliam, 1998), and (5) *Role of Assessing* (Bell & Cowie, 2001; Stiggins, 2005; Stigler & Hiebert, 1999). See Appendix A for the EQUIP descriptive rubric used to measure the Assessment indicators.

Method

Participants.

Twenty-two middle school math and science teachers from two highly diverse middle schools from the state's largest school district participated in a PD experience focused on increasing the quantity and quality of inquiry facilitated in their classrooms. As an indication of the poverty level of the student body, School A had a 33.3% free-and-reduced lunch population while School B had 57.7%. Twenty-one teachers (10 from School A and 11 from School B), ranging from 1-30 years of experience with a median of 5 years, continued for the entire year with one teacher dropping midyear. All 22 teachers attended a two-week summer institute during which they interacted with the 4E x 2 Instructional Model and became familiar with EQUIP. Subsequently, four follow-up sessions were held during the academic year. During these sessions, EQUIP was used to guide reflections and discussions held between PD facilitators and participants. Discussions focused on areas where growth had occurred and areas where challenges still remained regarding the implementation of inquiry-based instruction. In conjunction with this, four PD facilitators, including a science education researcher, a mathematics education researcher, and two doctoral students studying in the Curriculum and Instruction program area, conducted 102 observations and provided additional individual support throughout the year in the form of assistance with planning, debriefing after lessons, providing instruction while the teacher observed, and co-teaching.

Data collection and analysis.

This mixed methods study follows a triangulation convergence model (Creswell, 2008). In this model, quantitative and qualitative data are collected simultaneously, analyzed separately, and then mixed at the interpretation level. This model allowed for a thorough cross-comparison of findings from observational data and teacher self-reports, which provided a comprehensive view of teachers' growth in instructional practice.

Teachers' perceptions of growth and challenges in inquiry practices.

Teachers reflected on their inquiry practices using the indicators and descriptive rubric of EQUIP as a guide. These reflections occurred at follow-up meetings held bi-monthly during the academic year. Teachers indicated the two areas in which they felt they had experienced the greatest growth and the two areas where they faced their greatest challenges in relation to their inquiry practices. They justified their responses by providing detailed explanations and examples from their own teaching. These teacher reflections, collected during the third quarter of the academic year, were coded according to their alignment with indicators on EQUIP and analyzed for common themes across areas of perceived growth and challenge.

Observed growth and challenges in inquiry practices.

PD facilitators conducted observations in math and science classrooms using EQUIP's descriptive rubrics which detail levels of pre-inquiry, developing inquiry, proficient inquiry, and exemplary inquiry for each indicator (see Appendix A). (Although EQUIP also measures specific aspects of time usage, such as student attention levels and cognitive activity displayed by students at 5-minute increments, these data were not used in the analysis of this study.) Detailed field notes were also taken throughout the observations, providing a source of qualitative data to triangulate with quantitative ratings from EQUIP. The data analyzed for this manuscript are from 102 observations made during the academic year. The first and second quarter data comprise the pre-measure, and collectively the third and fourth quarter data create the post comparison group. Since the pre-data are gathered during the intervention year, we expect that growth will be somewhat understated because teachers begin growing after the summer intervention—not just between terms.

Observers used Microsoft Access to record EQUIP ratings, which were later transferred to SPSS for statistical analysis. A one-tailed, independent samples t-test (N = 102, df = 100) was conducted to identify statistically significant areas of teacher growth in relation to the indicators and scales of EQUIP. The two means being compared were (1) the observational data collected during the Fall Semester and (2) observational data collected during the Spring Semester. Finally, composite scores were determined based on the essence of performance in that category or scale. Even though growth in teacher practice is an important component of this study, it is also important to acknowledge the areas in which teachers are experiencing challenges. As such, a frequency analysis was conducted of each indicator studied. Specifically, of the 102 total observations made, we determined what percentage fell at or above the Proficient benchmark and what percentage fell below. The observed EQUIP measures were then compared with the teachers' perceptions in order to determine areas of alignment and divergence.

EQUIP validity and reliability issues.

EQUIP indicators were examined for internal consistency using Cronbach's Alpha (α) for all 102 class observations. The α-value ranged from .880-.889, demonstrating strong internal consistency. In addition, issues such as content and face validity have been previously addressed and substantiated (Marshall, et al., 2010). Further, we conducted 16 paired observations to analyze inter-rater reliability, via Cohen's Kappa (κ). The κ scores averaged between .55 and .62 for the four constructs. For these 16 paired observations, the coefficient of determination, r^2, was .856. The r^2 value indicates a more collective view of agreement between the raters. Specifically, 85.6% of Observer B's assessment is explained by Observer A's assessment and vice versa.

Results

Teacher growth: Perceived and observed.

Teachers' perceptions of growth in inquiry practice were measured using open-ended self-reports. With EQUIP indicators to guide them, teachers identified and then justified with evidence the areas of their teaching practice where the greatest personal growth occurred for the current school year. Nineteen of the 22 study participants were present to complete a self-report. These teachers reported greatest growth in the following areas: Instruction (16 responses), Discourse (10 responses), Assessment (8 responses), and Curriculum (2 responses). Note that participants could identify the top two areas of growth.

In addition to teachers' self-reported data, 102 full class observations were conducted using EQUIP. Our statistical analysis revealed that the means increased significantly on 12 of the 19 indicators, 3 of the 4 constructs, and for the overall lesson composite (see Table 1). The maximum score possible for the indicators and composites was four and the minimum was one. The greatest observed growth, noted by pre-/post-difference, was seen on the Instruction and Curriculum composites.

Since teachers' perceptions guided this analysis, it is appropriate to further analyze the Instruction construct data to better understand how this growth may have occurred. The observational data aligns with teachers' perceptions because growth was observed in five of six measurements relating to instructional issues (the five indicators and the composite score). Specifically, 15 of 16 teacher responses relating to Instruction occurred in three indicators: (1) *Order of Instruction,* (2) *Teacher Role,* and (3) *Student Role*. The data, primarily qualitative, surrounding teacher perceptions and observed growth for these indicators are detailed below.

Order of instruction.

Five teachers identified *Order of Instruction* as the area of greatest perceived growth in inquiry practice. Central to *Order of Instruction* is the concept of students exploring and interacting with math and science content before explanation occurs. For example, a seventh grade mathematics teacher reflected on his growth because he perceived that he had made significant strides in "asking students to explore concepts before receiving explanation, then having students explain with help from me when needed."

Even though teachers cited *Order of Instruction* as an area of significant growth, a statistical difference was not seen between the early fall versus late fall observations for the group as a whole on the *Order of Instruction* indicator. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Teacher asked students to explore before explanation. Teacher and student explained." The Proficient benchmark was achieved in 46 out of the 102 total lessons observations (45.1%). An example of a lesson that earned a Proficient rating for *Order of Instruction* involved a sixth grade science class that simultaneously investigated simple machines and technological design. After a brief overarching discussion on everyday objects that use levers and pulleys, students were challenged to "design an Egyptian pyramid for Queen Nefretiti [the teacher dressed in Egyptian attire] that would lift bricks [scaled versions of the huge stone blocks] with the miscellaneous materials provided in a "simple machines box." The initial discussion helped to focus student thinking while intentionally avoiding modeling the solution for the students who would be presenting their designs and solutions in subsequent classes.

Of the 102 total lessons observed, 56 (54.9%) did not achieve the Proficient benchmark for the *Order of Instruction* indicator. An example of such a lesson was observed in a seventh grade math class that was studying the relationship between fractions and decimals. The teacher began the lesson by presenting procedures for converting fractions to decimals and then followed with guided practice. There was no opportunity for students to explore the relationship between fractions and decimals, and as noted in the observational notes, "The lecture, guided practice, and group practice were completely prescribed by the textbook." Students worked with mathematical content only after a complete explanation had been given. Even then,

Table 1: Pre and post comparison of growth of teacher performance

Indicator	M-pre (N=58)	SD-pre	M-post (N=44)	SD-post	Change in Mean	t-value
I1—Instructional Strategies	2.41	.726	2.71	.649	0.30	2.181*
I2—Order of Instruction	2.11	.868	2.33	.998	0.22	1. 133
I3—Teacher Role	2.34	.939	2.84	.812	0.50	2.901**
I4—Student Role	2.41	.871	2.76	.802	0.35	2.101*
I5—Knowledge Acquisition	2.32	.740	2.64	.667	0.32	2.286*
Instruction Composite	**2.30**	**.734**	**2.71**	**.749**	**0.41**	**2.771****
D1—Questioning Level	2.18	.922	2.48	.822	0.30	1.738
D2—Complexity of Questions	1.95	.834	2.40	.748	0.45	2.813**
D3—Questioning Ecology	2.09	.910	2.22	.956	0.13	.712
D4—Communication Pattern	1.98	.731	2.29	.795	0.31	2.057*
D5—Classroom Interactions	1.98	.792	2.07	.856	0.09	.553
Discourse Composite	2.08	.667	2.32	.796	0.24	1.596
A1—Prior Knowledge	2.14	.765	2.16	.875	0.02	.113
A2—Conceptual Development	2.34	.608	2.53	.627	0.19	1.564
A3—Student Reflection	1.93	.846	2.10	.892	0.17	.984
A4—Assessment Type	2.18	.691	2.28	.643	0.10	.708
A5—Role of Assessing	2.02	.731	2.31	.706	0.29	2.007*
Assessment Composite†	**2.09**	**.563**	**2.31**	**.681**	**0.22**	**1.780***
C1—Content Depth†	2.05	.680	2.52	.755	0.47	3.307***
C2—Learner Centrality	2.09	.772	2.52	.800	0.43	2.706**
C3—Integration of Content & Investigation	2.16	.861	2.55	.902	0.39	2.220*
C4—Organization & Recording Information	1.95	.714	2.24	.844	0.29	1.814*
Curriculum Composite†	**1.95**	**.608**	**2.48**	**.800**	**0.53**	**3.788*****
Overall Lesson Composite	**2.36**	**.574**	**2.59**	**.593**	**0.23**	**1.902***

*** $p < .05$, ** $p < .01$, ***$p < .001$. †Unequal variances assumed using Levene's Test.**

their work was in a perfunctory, rote manner.

Teacher role.

Three teachers identified *Teacher Role* as an area of perceived growth in inquiry practice. Central to *Teacher Role* is the concept of the teacher acting as a facilitator for the student investigation and during the learning of math and science concepts. A sixth grade science teacher noted, "I enjoy my role as navigator. I only know where I want them to end up. The trip itself is very exciting and very individual."

EQUIP observational data showed more growth and the highest post mean in the *Teacher Role* indicator when compared to the other 18 indicators (see Table 1). Further, 61 of the 102 total lessons observed (59.8%) achieved a rating of Proficient or above. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Teacher frequently acted as facilitator." An example of a lesson that exceeded Proficient, scoring Exemplary, on *Teacher Role* included an eighth grade math class that was studying direct and inverse relationships. During the lesson, students explored these numerical relationships using a variety of strategies that included constructing a scatter plot. It was noted during the observation that the teacher "never gave the students the relationship. Instead, she guided them with questions…to help them discover the inverse relationship." In this instance, the teacher acted as a facilitator, guiding students to construct their meaning of inverse relationships by engaging them in a discovery learning experience.

For the *Teacher Role* indicator, 41 (40.2%) of the 102 total observed lessons earned below Proficient. An example of such a lesson occurred in a sixth grade math class where the teacher was a "giver of knowledge," as opposed to a facilitator. This teacher predominantly lectured to cover content, giving extensive notes on procedures for changing decimals into percents and working examples for students. During this math lesson, the teacher gave students explicit instructions on the procedures to use in order to convert between decimals, fractions, and percents. This more controlling teacher role did not allow students the opportunity to discover or even explore the many relationships between decimals, fractions, and percents, but rather presented these relationships as rules.

Student role.

Seven teachers indicated *Student Role* as their greatest area of perceived growth. Central to *Student Role* is the view that the students need to be actively engaged learners throughout the lesson. A sixth grade science teacher noted that she "can engage students more than [she] could last year to hook them in and get them interested in a topic."

EQUIP observational data indicated significant growth in the *Student Role* indicator. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Students were active as learners (involved in discussions, investigations, or activities, but not consistently and clearly focused)." Lessons met or exceeded this benchmark in 60 of the 102 observed lessons (58.8%). An example of a lesson that exceeded Proficient, earning Exemplary, occurred during a sixth grade science classroom that focused on the process skills of inference and observation. To begin the unit on technological design, students were given a tray of materials and presented with the following tasks: (1) make observations about these materials and (2) develop questions related to the possible uses of these materials. This lesson represented an inquiry investigation "with multiple opportunities for students to actively engage with scientific questions." Throughout the lesson, students were "creating and testing scientific questions" and were also given the opportunity to "examine alternate explanations" that were presented by their peers.

Lessons that scored lower than the Proficient benchmark on the *Student Role* indicator occurred in 42 (41.2%) of the 102 total observed lessons. An example of such a lesson occurred in a sixth grade science class that was studying independent and dependent variables. During this largely teacher-centered lesson we noted, "Students were active only when answering direct questions posed by the teacher." These questions tended to be low-level probes that followed a didactic questioning patterning, requiring minimal student thought or engagement. In this way, students exhibited short bursts of minimal engagement, followed by longer periods of passivity as they responded to teacher questions in this teacher-controlled environment. During this lesson, students had minimal opportunities to engage with scientific content, providing solicited responses to knowledge-level questions.

Teacher challenges: Perceived and observed

Using EQUIP indicators as a guide, teachers also indicated the two areas they perceived to be of greatest challenge in their inquiry-based practice. Teacher responses, grouped by indicator and scale, were quantified to show frequency of occurrence. Their collective responses showed they perceived that their greatest challenges in leading effective inquiry teaching practice lay in Assessment (16 responses), while the other areas also presented some challenge: Curriculum (9 responses), Instruction (6 responses), and Discourse (5 responses). Though it was not the area of smallest growth, the post-measures revealed that Assessment had the lowest overall mean. This fact, coupled with the teacher self-reports and observational data, confirmed that assessment issues could be considered the area of greatest challenge for the teachers. Although a significant difference in teacher growth was noted for the Assessment composite, only one of the five Assessment indicators, *Role of Assessment*, was significant.

Student reflection.

Five teachers responded that *Student Reflection*, one of the indicators for Assessment, was the area where they still felt the greatest challenge. Central to *Student Reflection* are students' explicit reflections on their understanding of math and science concepts. For example, a sixth grade science teacher noted that she "isn't comfortable with asking students to do reflections" and "hasn't been satisfied with the results" when she does ask students to reflect.

EQUIP observational data indicated no significant growth in the *Student Reflection* indicator, and only 34 of the 102 total observed lessons (33.3%) achieved Proficient or above. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Teacher explicitly encouraged students to reflect on their learning at an *understanding* level." All noted instances of *Student Reflection* involved student journaling or responses to reflective prompts. For example, one sixth grade teacher asked her students to respond to the following reflective prompt in their science journals: "Why is it important to have detailed materials and procedures when writing up a lab?"

Observed lessons earned below Proficient on the *Student Reflection* indictor in 68 of the 102 total observations (66.7%). Such an instance was noted during a sixth grade science class where "the teacher questioned students rapidly, giving them no time to reflect on the questions or their own understanding of the content." This below Proficient rating was earned when students were not provided an adequate and explicit opportunity to reflect on their own learning.

Assessment type.

Five teachers cited *Assessment Type* as their greatest perceived challenge in inquiry practice. *Assessment Type* refers to the teachers' use of formal and informal assessments in math and science classrooms. An eighth grade science teacher expressed her need for help in developing "more informal rubrics" and a desire to do more than give "true/false, fill-in-the-blank, and multiple choice assessments." Another teacher expressed her frustration that she could not seem to "get away from only using the tests."

As with *Student Reflection*, EQUIP observational data indicated no significant growth in the *Assessment Type* indicator, and only 33 out of the 102 total observations for this indicator (32.3%) earned Proficient or higher. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Formal and informal assessments used both factual, discrete knowledge and authentic measures." An example of a lesson that exceeded the Proficient benchmark on the *Assessment Type* indicator occurred in the context of a sixth grade science project entitled *Invention Convention*. Embedded within the technology and design unit, the *Invention Convention* provided students an opportunity to design a product, test its performance, and then make refinements in order to improve the product's performance. Numerous authentic assessments were noted throughout this project including the culminating project that was displayed for and discussed with judges at a regional competition. This authentic assessment allowed the teacher to examine the students' work across the entire design process and provided a complete picture of the students' ability to apply design principles in a real world context.

Of the 102 total lessons observed, 69 (67.7%) scored below Proficient on *Assessment Type*. Examples of such lessons included strictly short-answer assessments that focused on students' recall of factual, discrete knowledge and on which students responded to completion exercises and low-level didactic teacher questioning. For example, an eighth grade science teacher assessed students' knowledge of plate tectonics using a multiple choice quiz containing items written only at the knowledge level. This assessment required students only to recall basic facts relating to plate tectonics such as the following question: *Which type of boundary occurs at a place where plates are moving apart? (Answer: C, Divergent)*. Multiple-choice questions such as this provided evidence only of low-level factual knowledge; there was no evidence of the students' depth of understanding or clues about misconceptions that they might hold. In this case, a correct answer may even suggest that the student understood that divergent means to move apart, but still possessed virtually no conceptual knowledge about plate tectonics.

Role of assessing.

Four teachers cited *Role of Assessing* as their greatest perceived challenge in inquiry practice. *Role of Assessing* refers to the way in which teachers gather information about student understanding and how they use this information. A sixth grade science teacher noted that she would "really love to be more adept at formative assessment" but reported that currently, "I feel like I don't do it well."

Role of Assessing was the only indicator on the Assessment scale in which teachers showed significant growth on EQUIP measures, yet only 32 of the 102 total observations (31.4%) achieved a rating of Proficient or higher. A Proficient rating (Level 3) on EQUIP is defined by the following descriptive rubric: "Teacher solicited explanations from students to assess understanding and then adjusted instruction accordingly." One such example occurred during a lesson in a sixth grade science classroom during a pre-assessment on properties of atmospheric layers. During this activity, the teacher's assessment of students' prior knowledge about layers of the Earth's atmosphere led her to restructure her instructional plan to address student misconceptions that emerged during this pre-assessment. A post lesson conversation with the teacher confirmed that she did indeed alter her instructional plan. She added an introductory lesson to address students' inaccurate ideas about the relationship between the atmosphere and space prior to beginning the lesson she had originally planned. During the pre-assessment, many students demonstrated a belief that jet planes fly in space, beyond the boundary of Earth's atmosphere. Before beginning the original unit on the layers of the Earth's atmosphere, she challenged students to discuss the distinction between space and the atmosphere.

Of the 102 total lessons observed, 70 (68.7%) earned below the Proficient benchmark on *Role of Assessing*. An example of such a lesson occurred in a sixth grade mathematics lesson focused on multiplication of fractions. While students were working through a section of guided practice, one student voiced the following statement: "You can't do that [multiply] because you have to have a common denominator before you can multiply fractions." Instead of addressing this student misconception and using it in a formative way to guide instruction, the teacher dismissed the student's comment and began a new problem.

Discussion

EQUIP, a protocol designed to guide the evaluation of the quality of inquiry being facilitated, provides a common tool for conversations between teachers and PD facilitators (Marshall, Horton, & White, 2009). Just as PD facilitators use this protocol to measure teacher inquiry practices, teachers can also use the instrument to reflect upon their practices, determine their strengths and weaknesses, and select specific areas in which they would like to improve. These multiple purposes of EQUIP provide PD facilitators a formative assessment tool for adjusting the progression of a sustained PD experience as well as a point of discussion with teachers during classroom visits. Teacher perceptions of their growth/challenges and observed growth/challenges also provide data for refinement of subsequent PD experiences to best support the mission of increasing the quality of inquiry instruction in math and science classrooms.

All of EQUIP's 19 indicators and 5 composite scores (including the overall lesson composite) relative to inquiry-based instruction practice were developed based on a similar expectation that anything with a mean score of at least 2.5 is considered Proficient (Marshall, Horton, Smart, & Llewellyn, 2008; Marshall, Horton, & White, 2009). With that standard in mind, five of the six Instruction measures averaged Proficient or higher in the post observations, but only one of the six Assessment indicators averaged Proficient or higher (see Table 1). Since teachers consistently perceived the indicators associated with Assessment to be the area of greatest challenge regarding implementation and facilitation of inquiry-based instruction, it was not surprising to see smaller relative means when compared to Instruction indicators. The discussion presented below centers on the Instruction construct, the area where teachers perceived the most growth, and around the Assessment construct, the area where the greatest challenge was experienced. The teacher growth specifically focuses on the *Order of Instruction, Teacher Role*, and *Student Role* indicators, and the greatest challenges experienced by teachers focuses on *Student Reflection, Assessment Type*, and *Role of Assessing* indicators.

Teacher growth: Perceived, observed, and facilitated

Allowing students to explore concepts before students or the teacher give a formal explanation was a central focus of the entire PD experience. Specifically, aspects associated with *Order of Instruction* are among the central tenets of the 4E x 2 Instructional Model that was led during the summer, during follow-up meetings, and during discussions held before or after classroom observations. Additionally, teachers learned through modeling and discussion how to allow students to play an active role in the explanation phase. Aspects relevant to *Order of Instruction* were modeled for teachers and were central to the lessons that they developed during the summer and implemented during the academic year. However, while teachers have a perception of growth, their current scores show that they still average below the Proficient benchmark and the growth that was observed was not statistically significant (see Table 1, page 17).

The issues surrounding two other Instruction indicators, *Teacher Role* and *Student Role*, appear to be slightly different. Several teachers perceived growth in these areas and the post comparison averages were the two highest scores of all the indicators, and they were above the benchmark for Proficient. Ironically, more time was devoted to issues surrounding *Order of Instruction* than either *Teacher Role* or *Student Role*—though these areas are clearly related. Regardless, considerable time was devoted during the PD experience to help teachers begin to transition toward being a facilitator of learning instead of a teller of facts. This was modeled by having participants experience inquiry through a student's perspective first, and then by constructing lessons that would engage students in active ways, draw upon prior knowledge, and build new knowledge through shared experiences.

Teacher challenges: Perceived, observed, and facilitated

The lower growth seen in the Assessment indicators and composite could be due to several factors: (1) perhaps teachers made changes after the summer intervention but prior to the school year observations, (2) perhaps this is an area where teachers were already proficient, or (3) perhaps the changes to the instructional framework and curricular framework of instruction precede a detectable change in the Assessment indicators relative to inquiry instructional practice. If option one is true, then it seems that the findings would underreport the actual improvements made in instruction because some teachers will have made the changes before the instructional year began. Option two is not true since pre- and post-means for Assessment indicators were generally much lower than in the other three areas (see Table 1). Perhaps option three provides the best understanding. Many conversations and initiatives (Marzano, 2006; Wiggins & McTighe, 1998), including the one facilitated by the featured PD experience that used the 4E x 2 Instructional Model, include assessment as a critical piece of effective instructional practice. There appears to be, however, a large difference between the theory that supports the importance of assessments, particularly formative assessments, and the actual practice that teachers facilitate.

During the summer PD experience and during follow-up meetings and interactions, formative assessments were a continual and critical portion of what was modeled, experienced, and discussed. At every inquiry instruction phase that was modeled, formative assessments, including diagnostic, were explicit and integral to the experience. These experiences included: (1) identifying and working with student misconceptions, (2) scaffolding learning throughout an entire investigation, (3) developing appropriate and authentic diagnostic, formative, and summative assessments, (4) facilitating higher-order thinking, and (5) adjusting instruction based on data from students. Despite the efforts and time devoted during the PD experience to these topics, the reality is that teachers still, on average, are not meeting the Proficient benchmark on the indicators and composite related to Assessment.

Perhaps these assessment issues are trailing factors, so as teachers become more Proficient in the instructional and curricular aspects, Proficiency in assessment issues will soon follow. If true, this would suggest that participants who join a second year PD experience would likely be ready and able to begin making such a transition in their practice.

In summary, based on our observations and on teachers' perceptions, PD participants saw considerable growth in the area of Instruction. Further, a little more than half of the lessons observed for a given Instruction indicator achieved the Proficient benchmark, and the mean for the post comparison for the Instruction composite and four of the five indicators earned above the Proficient benchmark. This is particularly encouraging since we conducted our observations when the teachers indicated they would be leading inquiry-based instruction as well as when they did not perceive inquiry to be the desired strategy.

In contrast, participants saw, and our observations confirmed, that their greatest challenge lay in Assessment. Only about one-third of the observed lessons earned a rating of Proficient or higher, and only one of five indicators averaged Proficient or above in the post comparison.

Implications for Teacher Practice and Professional Development

The findings from this study may be beneficial in understanding how the intervention of the PD experience led to improved facilitation of inquiry-based practices by middle school math and science teachers, but it also provides a clear indication of areas where teachers often struggle. The EQUIP was central to identifying the strengths and challenges that PD participants experienced.

Though sufficiently challenging in their own rights, Instruction may be easier for teachers to alter than Assessment. In particular, modifying specific patterns of instruction may be manageable for teachers when given sufficient experiences, time, and support. However, modifying assessment proved to be a greater hurdle, at least for the teachers in our study. One explanation that we initially considered was that assessment naturally follows the curriculum and accompanying instruction; consequently, we might expect assessment to lag behind transformations in these areas. This theme was evident even though we had spent considerable time, energy, and scaffolding of learning in our efforts to make formative assessments more prevalent in the classroom. Though it is still possible that teachers will show greater improvement in assessment during the second year of the program, it may be the case that altering assessment practices, and, specifically, integrating formative assessment throughout the entire instructional process, is simply more difficult.

To explore this idea, we will need to delve into teachers' preconceptions and misconceptions about assessment and give the teachers opportunities to explore different forms of assessment as they implement inquiry-based instruction. Perhaps changes will come more easily for teachers as the transformations in their instruction become more natural for them, but perhaps we will need to focus our efforts more specifically on techniques for assessing students formatively as instruction is provided.

The art and science of teaching is indeed complex. However, with the framework provided by the EQUIP and by teachers' own perceptions, we believe we can continue to help teachers improve the quality of their inquiry-based instruction.

References

Abell, S. K., & Lederman, N. G. (2007). *Handbook of research on science education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536-553.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347-364.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 9-21.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school (expanded edition)*. Washington, DC: National Academies Press.

Bybee, R. W., Taylor, J. A., Gardner, A., Scotter, P. V., Powell, J. C., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins, effectiveness, and applications (pp. 49). Colorado Springs: BSCS.

Chambers, S. K., & Andre, T. (1997). Gender, prior knowledge, interest and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching, 34*(2), 107-123.

Chiappetta, E. L., & Koballa, T. R. J. (2006). *Science instruction in the middle and secondary schools: Developing fundamental knowledge and skills for teaching* (6th ed.). Upper Saddle River, NJ: Pearson Perrill Prentice Hall.

Chinn, C. A., & Brewer, W. F. (1998). Theories of knowledge acquisition. In B. J. Fraser & K. Tobin (Eds.), *International Handbook of Science Education* (pp. 97-113). Great Britain: Kluwer Academic Publishers.

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. In R. B. Davis, C. A. Maher & N. Noddings (Eds.), *Constructivist Views of the Teaching and Learning of Mathematics* (pp. 125-146). Reston, VA: NCTM.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. London: Taylor & Francis Ltd.

Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching, 14*, 169.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29-63.

Llewellyn, D. (2007). *Inquiry within: Implementing inquiry-based science standards in grades 3-8* (2nd ed.). Thousand Oaks, CA: Corwin Press

Marshall, J. C. (2009). *The Creation, Validation, and Reliability Associated with the EQUIP (Electronic Quality of Inquiry Protocol): A Measure of Inquiry-Based Instruction*. Paper presented at the National Association of Researchers of Science Teaching Conference, Orange County, CA.

Marshall, J. C., Horton, B., & Smart, J. (2009). 4E x 2 Instructional Model: Uniting three learning constructs to improve praxis in science and mathematics classrooms. *Journal of Science Teacher Education, 20*(6), 501-516.

Marshall, J. C., Horton, B., Smart, J., & Llewellyn, D. (2008). EQUIP: Electronic Quality of Inquiry Protocol. Retrieved from www.clemson.edu/iim

Marshall, J. C., Horton, B., & White, C. (2009). EQUIPping teachers: A protocol to guide and improve inquiry-based instruction. *The Science Teacher, 76*(4), 46-53.

Marshall, J. C., Smart, J., & Horton, R. M. (2010). The Design and Validation of EQUIP: An Instrument to Assess Inquiry-Based Instruction. *International Journal of Science and Mathematics Education, 8*(2), 299-321.

Marzano, R. J. (2006). *Classroom assessment and grading that work*. Alexandria, VA: ASCD.

Mezirow, J. (1990). *Fostering critical reflection in adulthood. A guide to transformative and emancipatory learning*. San Francisco: Jossey-Bass.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead, UK: Open University Press.

National Board for Professional Teaching Standards. (2006). Making A Difference in Quality Teaching and Student Achievement Retrieved from http://www.nbpts.org/resources/research

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: NCTM.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (pp. 703-732). New York: Wiley.

Renner, J. W., Abraham, M. R., & Birnie, H. H. (1988). The necessity of each phase of the learning cycle in teaching high school physics. *Journal of Science Teacher Education, 25*, 39-58.

Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324-328.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

van Zee, E. H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching, 38*(2), 159-190.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118.

White, B. Y., & Frederiksen, J. R. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist, 40*(4), 211-223.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

Jeff C. Marshall, Ph.D., is an associate professor and director of Inquiry in Motion, Moore School of Education, Clemson University, 418G Tillman Hall, Clemson, SC 29634. Inquiries may be sent to marsha9@clemson.edu or view the website for further information: www.clemson.edu/iim

Julie Smart, Ph.D., is an assistant professor, Presbyterian College, 503 S. Broad St., Clinton, SC 29325.

Robert M. Horton, Ed.D., is a professor, Moore School of Education, Clemson University, 409B Tillman Hall, Clemson, SC 29634.

Acknowledgements: This work is supported by the National Science Foundation (Grant #DRL-0952160) and by South Carolina Center's of Excellence program.

Appendix A (Marshall, et al., 2008)

IV. Instructional Factors					
Construct Measured		*Pre-Inquiry (Level 1)*	*Developing Inquiry (2)*	*Proficient Inquiry (3)*	*Exemplary Inquiry (4)*
I1.	Instructional Strategies	Teacher predominantly lectured to cover content.	Teacher frequently lectured and/or used demonstrations to explain content. Activities were verification only.	Teacher occasionally lectured, but students were engaged in activities that helped develop conceptual understanding.	Teacher occasionally lectured, but students were engaged in investigations that promoted strong conceptual understanding.
I2.	Order of Instruction	Teacher explained concepts. Students either did not explore concepts or did so only after explanation.	Teacher asked students to explore concept before receiving explanation. Teacher explained.	Teacher asked students to explore before explanation. Teacher and students explained.	Teacher asked students to explore concept before explanation occurred. Though perhaps prompted by the teacher, students provided the explanation.
I3.	Teacher Role	Teacher was center of lesson; rarely acted as facilitator.	Teacher was center of lesson; occasionally acted as facilitator.	Teacher frequently acted as facilitator.	Teacher consistently and effectively acted as a facilitator.
I4.	Student Role	Students were consistently passive as learners (taking notes, practicing on their own).	Students were active to a small extent as learners (highly engaged for very brief moments or to a small extent throughout lesson).	Students were active as learners (involved in discussions, investigations, or activities, but not consistently and clearly focused).	Students were consistently and effectively active as learners (highly engaged at multiple points during lesson and clearly focused on the task).
I5.	Knowledge Acquisition	Student learning focused solely on mastery of facts, information, and/or rote processes.	Student learning focused on mastery of facts and process skills without much focus on understanding of content.	Student learning required application of concepts and process skills in new situations.	Student learning required depth of understanding to be demonstrated relating to content and process skills.

VI. Assessment Factors					
Construct Measured		*Pre-Inquiry (Level 1)*	*Developing Inquiry (2)*	*Proficient Inquiry (3)*	*Exemplary Inquiry (4)*
A1.	Prior Knowledge	Teacher did not assess student prior knowledge.	Teacher assessed student prior knowledge but did not modify instruction based on this knowledge.	Teacher assessed student prior knowledge and then partially modified instruction based on this knowledge.	Teacher assessed student prior knowledge and then modified instruction based on this knowledge.
A2.	Conceptual Development	Teacher encouraged learning by memorization and repetition.	Teacher encouraged product- or answer-focused learning activities that lacked critical thinking.	Teacher encouraged process-focused learning activities that required critical thinking.	Teacher encouraged process-focused learning activities that involved critical thinking that connected learning with other concepts.
A3.	Student Reflection	Teacher did not explicitly encourage students to reflect on their own learning.	Teacher explicitly encouraged students to reflect on their learning but only at a minimal knowledge level.	Teacher explicitly encouraged students to reflect on their learning at an understanding level.	Teacher consistently encouraged students to reflect on their learning at multiple times throughout the lesson; encouraged students to think at higher levels.
A4.	Assessment Type	Formal and informal assessments measured only factual, discrete knowledge.	Formal and informal assessments measured mostly factual, discrete knowledge.	Formal and informal assessments used both factual, discrete knowledge and authentic measures.	Formal and informal assessment methods consistently and effectively used authentic measures.
A5.	Role of Assessing	Teacher solicited predetermined answers from students requiring little explanation or justification.	Teacher solicited information from students to assess understanding.	Teacher solicited explanations from students to assess understanding and then adjusted instruction accordingly.	Teacher frequently and effectively assessed student understanding and adjusted instruction accordingly; challenged evidence and claims made; encouraged curiosity and openness.