

ASK-IT Information

Project Director-Gloria Ferguson, ESD 112

Two year research project funded by WA State LASER

Partners:

- Six schools, 3 treatment and 3 control
- ESD 112, Vancouver WA
- Kathy Long-Assessment Developer for FOSS, Lawrence Hall of Science UC Berkeley
- Cathy Kennedy-Research Design & Evaluator

Hypothesis: If you use quality assessment data, and take appropriate action based on that data, student learning will improve.

The research “treatment”

- 1) The intentional use of ASK assessment tools to monitor & improve instruction
- 2) Grade-level teachers working together (PLC) on the implementation of ASK resources and best practices
- 3) Increased time in science by integrating with math and language arts

Research Project Data Sources

- Student achievement data
 - ASK pre/post data specific to science module
 - WASL scenarios
 - WASL scores
- Teacher practice
 - Pre/post teacher attitude survey
 - Focus group questions
 - Teacher logs
 - Classroom observation

ASK-IT Year 1 Report **August 15, 2008**

Theory: Teachers who (1) are trained in ASK Assessment, (2) participate in a PLC, and (3) increase time in science teaching improve their teaching practice and subsequently improve the learning of their students.

Treatment: Teachers who participated in all three tiers.

Control: Teachers who participated in two or fewer tiers. [NOTE: Although originally recruited as treatment teachers, Wallace teachers are considered part of the control group for these analyses since they only partially implemented the treatment.]

Findings

Hypothesis 1: Teachers who participate in the treatment *demonstrate more use of formative assessment for classroom decision making* than teachers in the control group.

In this first year of research, we found that ASK-IT teachers use formative assessment more often than control group teachers, and use it to learn what their students need and to make classroom decisions more often. We also found that these teachers used formative assessment in their classes as soon as they had completed the formative assessment workshop. In addition, despite our best attempts to match districts, schools and teachers for our treatment and control groups, teachers in the treatment group had a slightly more positive belief about the value of regular formative assessment than did teachers in the control group. Finally, we found that teachers who participated in the treatment group spent significantly more time on science in their classes, and were more likely to complete the entire FOSS module. Teachers in the control group spent far less time in science overall, and were unlikely to finish the entire module.

In focus group discussions, we found that ASK-IT teachers learned that they could have high expectations for their students. As one teacher put it, “I believe students are more capable of higher level thinking than what I had once thought.” We also found that teachers use formative assessment as a way of keeping themselves, as well as their students, accountable for student outcomes. The following are excerpts from focus group discussions:

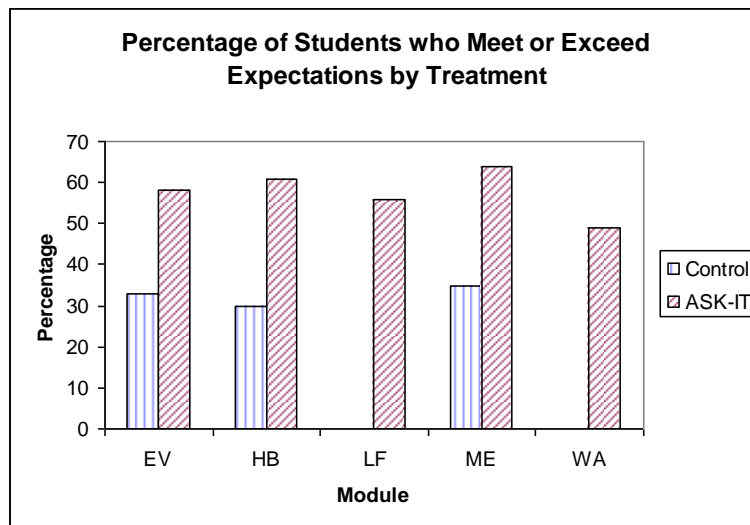
- “By being involved in ASK-IT, I was holding myself and my students more accountable. In the past, I did not take enough time to go through the notebooks and did not re-teach missed concepts. [In ASK-IT,] the students were more involved with science writing and were allowed to share, discuss, and build off of each others’ ideas.”
- “I’ve become more aware of how to help students increase their abilities to express their thinking/understanding (either verbally or in written form) at a much deeper level of cognition, supporting their ideas with facts and data, as opposed to ‘just because.’ Also, [I learned] that kids are very capable of taking their idea and expressing it in scientific jargon. I’m holding them to a higher level of expectation.”

- “The journals have helped me become a better science teacher! After viewing journal entries, I retaught and reviewed that lesson. Before, I would briefly restate and move on. Also, the students were held more accountable. If they missed the point of the questions, they were asked to use another color pencil/pen and add to their answer to improve the quality of their response. The kids weren’t threatened by this and by the end of the module, their writing science skills greatly improved!”

We also found that teachers were using what they had learned about good formative assessment practice in their classes. One teacher mentioned how when he moved away from always assigning grades to student work, students became more responsible for self-monitoring. He told us, “Not having grades on their I-checks forced students to evaluate their answers and to think about what they’ve learned.” Another said, “Notebooks and the opportunity to journal allows students to show their learning in a variety of ways. The ability to explain the rationale in their hypothesis leads to deeper understanding. The use of the notebooks helps move the focus from just ‘doing’ the investigations to ‘understanding’ the investigation concept.”

Hypothesis 2: Students of teachers who participate in the treatment will have *greater learning gains* than students of teachers who did not participate.

In every module for which we collected data (Environments, Human Body, Landforms, Magnetism & Electricity, and Water) students in the treatment group outperformed students in the control group and the differences were statistically significant when we controlled for prior knowledge (using pretest scores). Each FOSS module defines four categories, or levels, of student performance: Needs Instruction, Approaches Expectations, Meets Expectations, and Exceeds Expectations. Students in the two groups performed similarly on all of the pretests (called Surveys in the FOSS curriculum), but students in the treatment group were far more likely to meet or exceed expectations on the post test than students in the control group. On average, students in the control group did not meet expectations on the post test in any of the content areas.



We also found some differences in the performances of students in different classes within the treatment group. In the Year 2 study we will be able to correlate these differences with teacher

attitudes (e.g., confidence in teaching science, belief in the value of formative assessment) and practices (e.g., feedback provided to students, length of time spent on science).

Details of the Analyses

Hypothesis 1: Teachers who participate in the treatment *demonstrate more use of formative assessment for classroom decision making* than teachers in the control group.

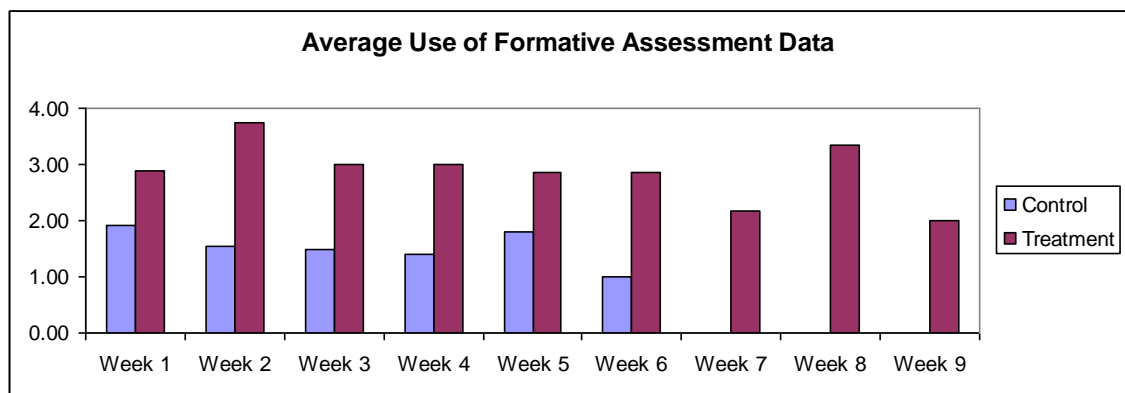
Result 1 – Teachers in the treatment group were more likely than teachers in the control group to use formative assessment to inform teaching on a weekly basis. In weekly logs of classroom activity, teachers could indicate when they reviewed student work, what they learned, and how they responded to student needs. 12 treatment group teachers and 11 control group teachers completed weekly logs during the Spring 2008 term. These were encoded as scores of 1, 2, 3 or 4 for each week, as follows:

- 1 = no evidence of review of student work
- 2 = evidence of review
- 3 = evidence of learning from review
- 4 = evidence of action resulting from review

An average level of use of formative assessment was then computed for each teacher (averages ranging from 1 to 4 based on the categories).

Teachers in the treatment group were more likely to state that they had learned something about their students' needs from evaluating classroom work (weekly average = 2.85), while teachers in the control group showed evidence of having reviewed student work about half the time (weekly average = 1.52), and rarely mentioned that they had learned anything or changed their teaching plans based upon that review. This difference was statistically significant at the $\alpha = .05$ level ($p = .001$).

Result 2 - In our analysis of week-by-week use of formative assessment, we found that teachers in the two groups differed from the very beginning of the study, with teachers in the treatment group tending to use formative assessment data more than teachers in the control group, as illustrated in the figure below. The blue columns show the average use of formative assessment for teachers in the control group for each week the module was taught (1=not used, 2=reviewed student work, 3=learned something from review of student work, 4=took action based on formative assessment data). All of the control group teachers completed the module on or before Week 6. The red columns show the average use of formative assessment for the teachers in the treatment group.



Result 3 - Because teachers in the treatment group participated in a two-day workshop about formative assessment prior to starting the first FOSS module, we are not surprised that the weekly logs show immediate teacher differences. To account for any inherent differences in teachers in the two groups, we evaluated teachers' initial attitudes about formative assessment as reflected in their responses to items on the *Teacher Attitude Survey*, which was administered before the first workshop.

The following questions from the *Teacher Attitude Survey* provide insight of a teacher's belief about the value of formative assessment:

- d. Embedded assessment should be used in every lesson.
- e. Written work is a better indicator than oral conversation regarding student understanding.
- o. Formative assessment should provide individualized information about each student.
- p. Teachers need to provide time for students to reflect on their own thinking process.

We found that teachers in the treatment and control groups had similar attitudes on items e, o and p,¹ but differed in their endorsement of the statement "Embedded assessment should be used with every lesson." On average, teachers in the treatment group agreed with this statement while teachers in the control group were noncommittal. Although this one point difference in the 5-point Likert scale was statistically significant at the $\alpha = .05$ level ($p = .003$), the difference does not appear to indicate a strong bias for or against formative assessment. This finding is actually understandable. Given that schools were given the option of participating as a treatment school or a control school, and knew what would be required if they chose to be a treatment school, we would expect teachers who volunteer to be in the treatment group to believe that formative assessment is a valuable tool for teaching.

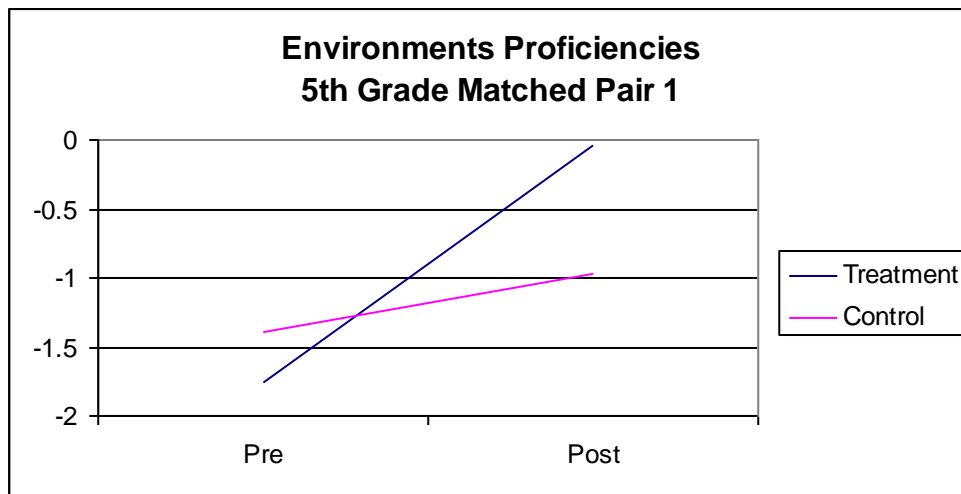
Result 4 – Even with this relatively small sample of 23 teachers who completed weekly logs, we found that the teachers who participated in the treatment were more likely than control group teachers to (1) complete teaching the FOSS module ($r = .694$, $p < .001$) and (2) use formative assessment data ($r = .501$, $p = .013$) to make classroom decisions. Only 8% of control group teachers completed all investigations of the module they were teaching in Spring 2008, compared with 50% of the treatment group teachers. On the other hand, 69% of control group teachers completed fewer than three investigations in their module, while no treatment group teachers completed so few investigations in their modules.

¹ They disagreed with the statement, "Written work is a better indicator than oral conversation regarding student understanding," they were noncommittal about whether "Formative assessment should provide individualized information about each student," and they strongly agreed that "Teachers need to provide time for students to reflect on their own thinking processes."

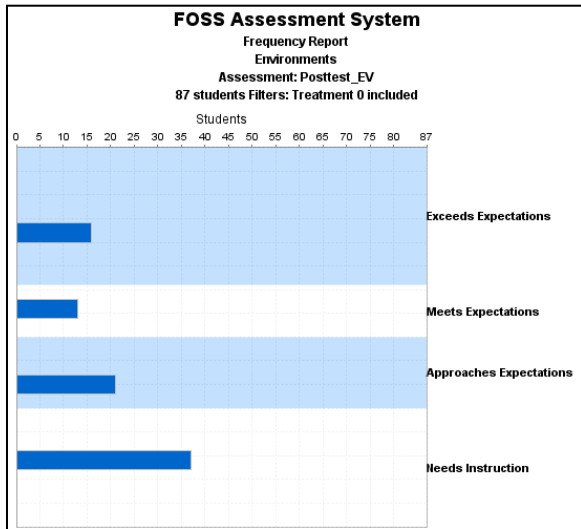
Hypothesis 2: Students of teachers who participate in the treatment will have *greater learning gains* than students of teachers who did not participate.

EV Environments Module – Includes data from paired schools 1T and 1C. Group averages for the pre-instruction survey and post test are shown in the table below. At the beginning of the module, students in both groups have statistically similar pretest scores ($p = .275$) placing students at the boundary between the “Needs Instruction” and “Approaches Expectations” levels on the construct. By the end of the module, on average, students in the control group have advanced to the “Approaches Expectations” level ($\mu = -0.974$, cut score = -1.5), while students in the treatment group have advanced to boundary of the “Meets Expectations” level ($\mu = -0.028$, cut score = 0.0). This difference is significant at the $\alpha = .05$ level ($p = .002$) and is illustrated in the chart below. Students in the treatment class (the black line) began with slightly less understanding of concepts in the Environments module than students in the control class, but by the end of the module they outperformed control group students on the post test.

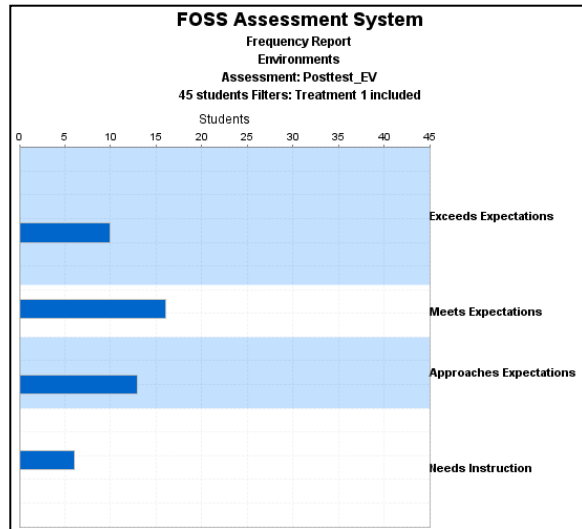
	Avg. Survey	Avg. Post
Overall:		
Treatment	-1.748 (50)	-.028 (45)
Control	-1.386 (43)	-.974 (87)



A more detailed breakdown of student performance on the post test is shown in the histogram chart in the figure below. The alternating horizontal bands of blue and white in the chart indicate different levels of understanding of the Environments module concepts: proficiencies in the lowest white band indicate that students need additional instruction, the next higher blue band indicates that students are approaching expected understanding, the next white band indicates students who meet expectations for the module, and the highest blue band shows proficiencies that indicate that students are exceeding expected performance in the module. At the end of the module, only 33% of control group students met or exceeded expectations, while 58% of treatment group students met or exceeded expectations.



Performance on Post test for Control group.

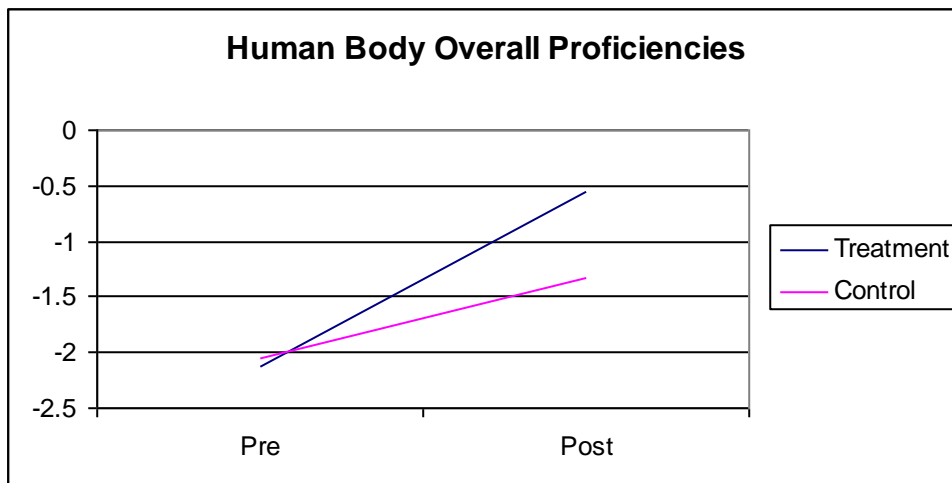


Performance on Post test for Treatment group.

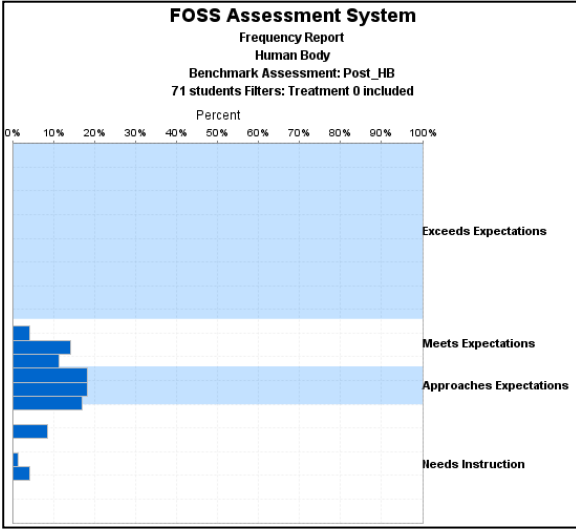
ANOVA analysis indicates no statistical significance of post test performance by class membership among students in the treatment group when we control for pretest performance ($p = .773$).

HB Human Body Module – Includes data from paired schools 1T and 1C. Group averages for the pre-instruction survey and post test are shown in the table below. At the beginning of the module, students in both groups have similar pretest scores ($p = .595$) placing students in the “Needs Instruction” category on the construct. By the end of the module, on average, students in the control group have advanced to the “Approaches Expectations” level ($\mu = -1.222$, cut score = -1.5), while students in the treatment group have advanced to the “Meets Expectations” level ($\mu = -0.558$, cut score = -0.7). This difference, which is illustrated in the chart below, is significant at the $\alpha = .05$ level ($p < .001$).

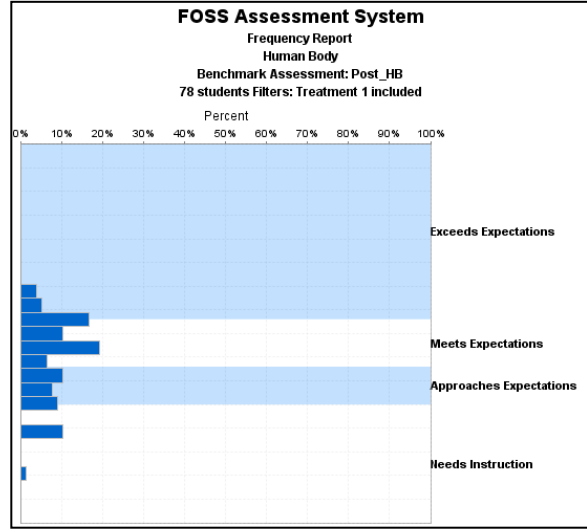
	Avg. Survey	Avg. Post
Overall:		
Treatment	-2.130 (72)	-0.558 (78)
Control	-2.047 (70)	-1.222 (70)



A more detailed breakdown of post test performance is shown in the figure below. At the end of the module, only 30% of control group students met expectations, while 61% of treatment group students met or exceeded expectations.



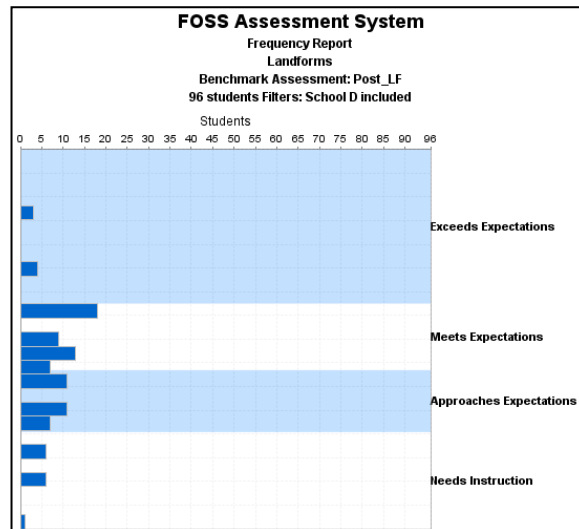
Performance on Post test for Control group.



Performance on Post test for Treatment group.

ANOVA analysis of post test performance by class indicates that among students in the treatment group, class membership had a significant influence on post test scores when we controlled for pretest performance ($p < .001$).

LF Landforms Module – Includes data from schools 2T and 3T. These are both treatment schools, so no treatment-control comparison could be made. On average, students started in the “Needs Instruction” level and finished in the “Meets Expectations” level ($\mu = -0.497$, cut score = -0.65). As shown in the figure below, at the end of the module 56% of treatment group students met or exceeded expectations.

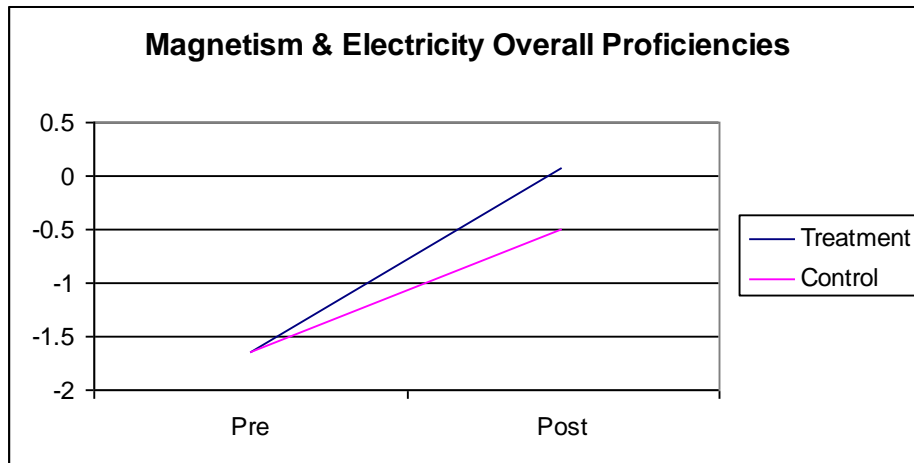


Performance on Post test for Treatment group.

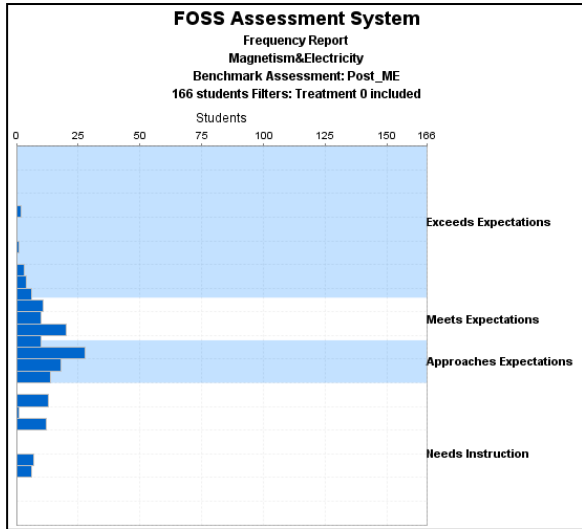
ANOVA analysis indicates no statistical significance of post test performance by class membership among students in the treatment group when we control for pretest performance ($p = .141$).

ME Magnetism & Electricity Module – Includes data from schools 1T, 1C, 2C, 4T, and 4C. Group averages for the pre-instruction survey and post test are shown in the table below. A linear regression of post test performance controlling for pretest performance indicates that membership in the treatment group is a statistically significant factor ($p = .001$). Students in the treatment group outperformed students in the control group. At the beginning of the module, students in both groups have similar pretest scores ($p = .567$) placing students, on average, in the “Needs Instruction” category on the construct. By the end of the module, on average, students in the control group have advanced to the “Approaches Expectations” level ($\mu = -0.424$, cut score = -1.0), while students in the treatment group have advanced to the “Meets Expectations” level ($\mu = 0.075$, cut score = -0.1). This performance difference is illustrated in the chart below.

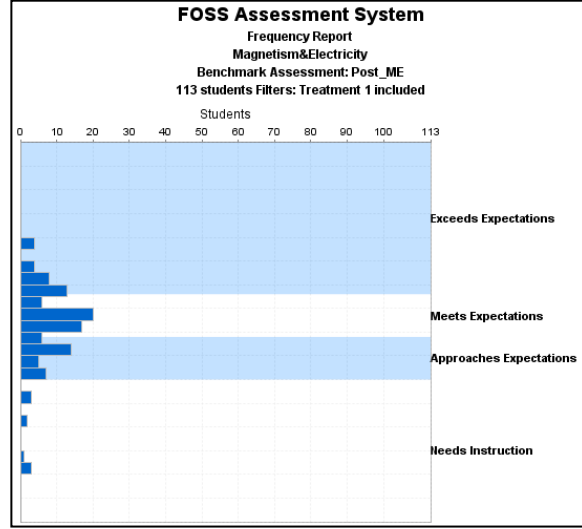
	Avg. Survey	Avg. Post
Overall:		
Treatment	-1.638 (115)	0.075 (113)
Control	-1.628 (201)	-0.424 (207)



A more detailed breakdown of student performance on the post test is shown in the figure below. At the end of the module, only 35% of control group students met or exceeded expectations, while 64% of treatment group students met or exceeded expectations.



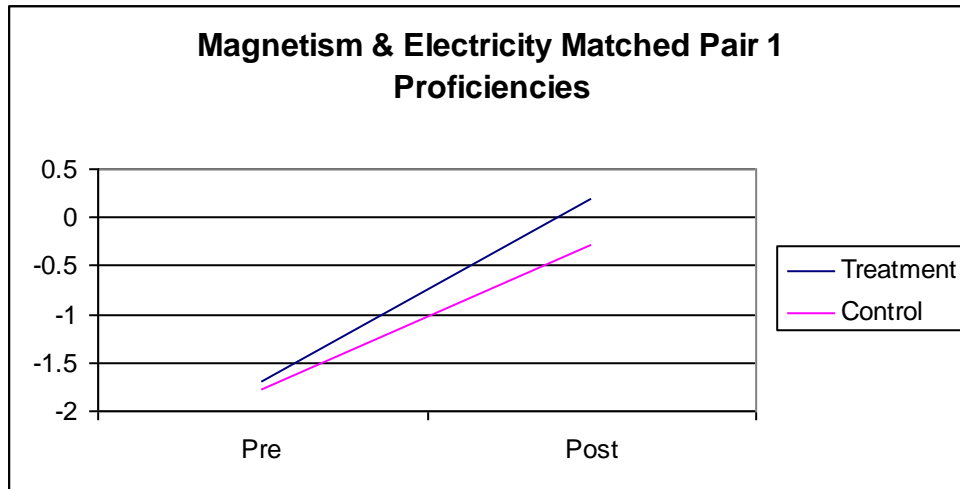
Performance on Post test for Control group.



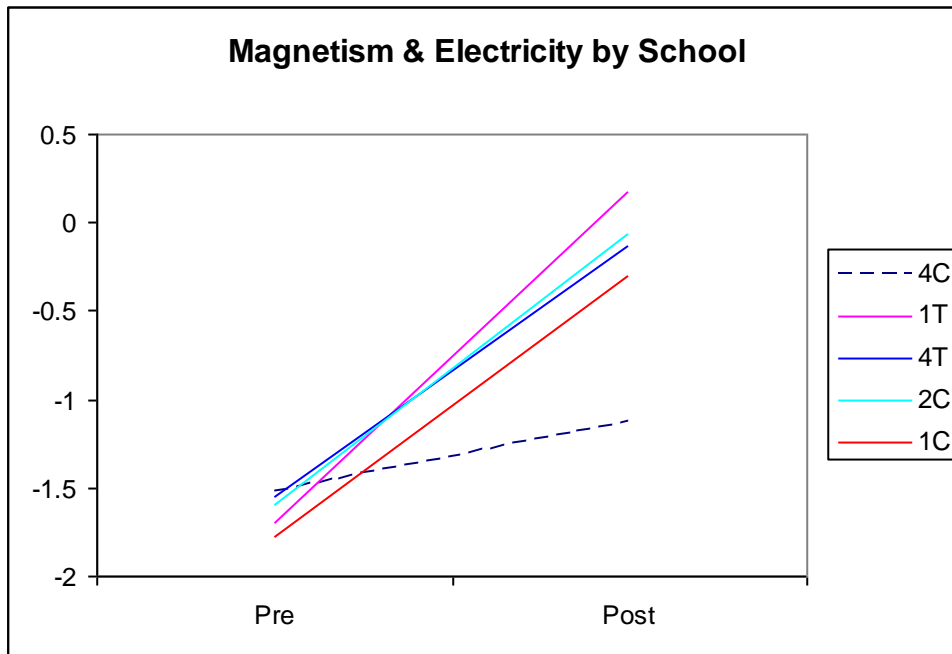
Performance on Post test for Treatment group.

An analysis of paired schools with similar demographics indicates the same pattern of performance, with similar pretest scores ($p = .569$) and different post test scores ($p = .002$), shown in the table below.

	Avg. Survey	Avg. Post
Pair 1:		
Treatment	-1.693 (73)	0.185 (72)
Control	-1.770 (69)	-0.291 (67)

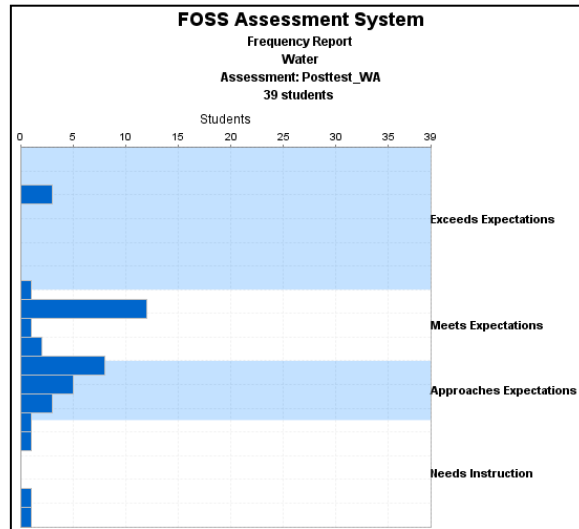


We also found significant differences in post test performance by school when we controlled for pretest performance, although the effect size was not great among schools overall. As shown in the chart below, students in school 4C exhibited smaller learning gains than students in the other schools, while students in school 1T exhibited the largest learning gains.



ANOVA analysis of post test performance by class indicates that among students in the treatment group, class membership had a significant influence on post test scores when we controlled for pretest performance ($p < .001$).

WA Water Module – Includes data from only one school (2T), so no treatment-control comparison could be made. On average, students started in the “Needs Instruction” level and finished in the “Meets Expectations” level ($\mu = -0.219$, cut score = -0.5). As shown in the figure below, at the end of the module 49% of treatment group students met or exceeded expectations.



Performance on Post test for Treatment group.

ANOVA analysis indicates no statistical significance of post test performance by class membership among students in the treatment group when we control for pretest performance ($p = .084$).