Abstract Title: Identifying and measuring factors related to student learning: the promise and pitfalls of teacher instructional logs

MSP Project Name: Assessing Teacher Learning About Science Teaching (ATLAST)

Author(s): P. Sean Smith and R. Keith Esch

Presenter(s): P. Sean Smith and R. Keith Esch

120 word summary:

Existing measures have produced inconsistent and weak evidence for claims about the relationships among teacher content knowledge, classroom practice, and student learning. The ATLAST project has developed pairs of assessments—for teachers and students—in three middle grades content areas: force and motion, flow of matter and energy in living systems, and plate tectonics. We used these instruments and classroom instruction information gleaned from teacher-generated logs to explore the relationships among teacher content knowledge, classroom instruction, and student learning. Our findings suggest the relationships among teacher knowledge, amount of instruction, and student learning depend on the science content; and that teacher instructional logs do not provide sufficient evidence to gauge the quality of instruction.

- Section 1: **Questions for dialogue at the MSP LNC**.
  A. What are the relationships among teacher content knowledge, classroom instruction, and student learning of science content?
  B. Can teacher logs be used to gauge the quality of classroom instruction (student opportunity to learn) reliably and validly; i.e., without observing instruction?

- Section 2: **Conceptual framework**. This section should include your project's definition of "student success" and an explanation about your project's design for measuring student success.
  A. Student Success

     In the context of the ATLAST project, student success is defined simply as the change in student scores on pre-instruction and post-instruction assessments.

  B. Context of the work within the STEM education literature and within your MSP Project

     Discussions of the relationship between science teacher characteristics and student learning tend to be theoretical and divisive. The lack of consensus is attributable in part to a thin empirical basis for the competing arguments. Despite the lack of consensus, there is broad agreement on one premise—teacher knowledge of disciplinary content directly and positively affects classroom practice and, ultimately, student learning. Interestingly, although this premise is logical, the empirical support is inconsistent, largely because of a lack of suitable measures. Studies rely mainly on proxies of teacher content knowledge, for example certification type (Goldhaber and Brewer, 2000), undergraduate major (Monk, 1994), and courses taken (Druva and Anderson, 1983). A recent study (Wallace, 2009) found a weak but consistent and positive relationship between professional development and student achievement in reading and mathematics. Few studies use direct measures of teacher content knowledge. Existing student measures tend to have weak psychometric properties, or they are broad (e.g., state-administered assessments, NAEP), further limiting the likelihood of finding relationships between teacher knowledge of particular content and student learning.

The myriad variables associated with classroom instruction further complicate the relationship. For instance, instructional materials may mask an otherwise strong relationship between teacher content knowledge and student learning.

ATLAST (an MSP-RETA) was funded to develop instruments related to teacher knowledge, classroom instruction, and student learning, all tightly aligned to the same, carefully defined strands of science content. We have developed pairs of psychometrically rigorous multiple-choice assessments—one for teachers and one for students—in three middle grades content areas: force and motion, flow of matter and energy in living systems, and plate tectonics. In addition, we have developed a daily instructional log and a reliable method for extracting information about classroom instruction from the log.

C. Claim(s) or hypothesis(es) examined in the work (anticipating that veteran projects will have claims, newer projects will have hypotheses)
   1. Hypothesis: Instructional materials are a powerful mediating influence in the relationship between teacher knowledge and student learning. (i.e. quality and appropriateness of instructional materials can confound the effect that teacher knowledge may have on student learning.)

   2. Claim: Teacher logs of classroom instruction are inadequate for clarifying the relationship between teacher content knowledge and student learning.

- <u>Section 3: **Explanatory framework**</u>. This section should describe what you are finding, or are set up to learn, about student success, and how it is informing, or will inform, your MSP work.

A. Evaluation, research design, data collection and analysis

Research Design

The primary purpose of this study was to explore the relationships among teacher content knowledge, classroom instruction variables, and student learning using the ATLAST measures. We recruited roughly 200 teachers nationally for each of two studies: one on force and motion and one on plate tectonics. To be eligible for the study, each teacher had to affirm that s/he taught a unit on the content of the study as part of their normal instruction. Teachers were asked not to alter their instruction in any way.

Data Collection

Once enrolled in the study, each teacher:

- Completed the relevant teacher assessment shortly before the unit of instruction;
- Administered the relevant student assessment just before and just after the unit; and
- Completed a daily web-based instructional log during the unit.

Teachers completed their assessment on-line. For the student assessments, we printed test booklets and answer sheets and sent these to the teachers, who gave the assessments to their students and returned them to us for scoring and analysis. Table 1 shows the internal reliabilities for the teacher and student assessments. In addition, the test-retest reliabilities for the Force and Motion and Plate Tectonics teacher assessments are 0.885 and 0.945, respectively.

Table 1.
Internal Reliabilities of Student and Teacher Assessments

Measure	IRT Reliability
Force and Motion	
Student	0.75
Teacher	0.85
Plate Tectonics	
Student	0.86
Teacher	0.86

Content validity of the assessment items was established through expert review and through cognitive interviews with teachers and students. Content experts judged the assessments to be accurate and adequate representations of the content domain. Interviews established that teachers and students interpreted the items as developers intended.

Teachers also completed a web-based instructional log for each day of the unit. The log was dynamic, with the information requested dependent on the instructional activities indicated by the teacher. For instance, if the teacher reported that students read a passage from the textbook, the log prompted the teacher for page numbers. Other information requested by the log included published and teacher-generated instructional materials used (if any), date of the lesson, length of the lesson (in minutes), and the specific purpose of the day's lesson.

Data collection took place from January to June 2009. There was substantial attrition at the teacher level from both studies, which we attribute to the burden associated with the teacher log. Of the teachers in the Force and Motion study, 79 completed all components, compared with 107 in the plate tectonics study.

Data Analysis

IRT scores were calculated for each student (pre- and post-unit) and teacher (pre-unit only). Means and standard deviations are shown in Table 2.

Table 2.
Mean IRT (Theta) Scores on Student and Teacher Assessments

Measure	Pre-unit				Post-unit			
Force and Motion	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Student (n=1689)	-0.00	0.64	-1.46	2.18	0.50	0.78	-1.46	2.62
Teacher (n=79)	0.25	0.82	-1.28	2.63	n/a	n/a	n/a	n/a
Plate Tectonics								
Student (n=2261)	0.14	0.85	-2.34	2.61	0.86	0.99	-2.19	2.61
Teacher (n=107)	0.20	0.83	-1.51	1.95	n/a	n/a	n/a	n/a

The only classroom instruction variable included in this analysis was amount of instruction, which proved a challenge to calculate. Although the log asked for the length of each lesson in minutes, we could not simply sum these, as we discovered that some instructional time was not aligned with the content that we assessed. We developed a reliable procedure for calculating the amount of aligned instructional time that involved analyzing each teacher's logs. Two pairs of researchers applied this procedure to a random sample of 30 teachers' logs in each content area. The interrater reliability for Force and Motion and Plate Tectonics were 0.929 and 0.826, respectively. The mean minutes of instruction for all teachers in each study is shown in Table 3.

Table 3.
Minutes of Instruction

Content Area	Minutes	S.D.	Min	Max
Force and Motion (n=79)	180.50	97.68	0.00	495.00
Plate Tectonics (n=107)	208.29	116.73	0.00	667.00

Data were analyzed using hierarchical linear modeling (HLM) to account for the variation shared by students in the same class. Student pre- and post-test scores were entered at the first level. Student demographic variables (race/ethnicity, gender, grade in school, English language learner status) were entered at the second level. Teacher variables (assessment score and hours of instruction) were entered at the third level.

In the Force and Motion study, amount of instruction by itself did *not* predict student learning. Teacher knowledge *was* a significant predictor of student learning, both by itself and with amount of instruction in the model. An increase of one standard deviation in teacher score translated to an increase of 0.12 standard deviations in student learning.

In Plate Tectonics, neither amount of instruction *nor* teacher knowledge predicted student learning, either by themselves or in combination.

In an attempt to clarify what is obviously a murky relationship between teacher content knowledge and student learning, we explored the possibility of mining the unit instructional logs for more information about student opportunity to learn.

Our approach was situated in a theory of teaching for understanding. Consistent with this theory, we adopted a variation of the learning cycle as the basis for analyzing student opportunity to learn:

1. Situating the learning;

2. Students expressing their initial ideas;
3. Students examining relevant phenomena;
4. Students making sense of phenomena; and
5. Students making sense of the targeted idea(s).

To construct a framework for analyzing unit instructional logs, we identified phenomena that support students' development of the targeted content for specified ideas of the two content areas. Distinctions were made between primary phenomena and evidentiary phenomena. <u>Evidentiary phenomena</u> are *observable*, naturally occurring events that provide evidence for the **primary** phenomena (e.g., <u>the distance between GPS ground stations on different tectonic plates changes</u>, providing evidence that **tectonic plates move**). HRI researchers (and content experts) used the identified phenomena as criteria for evaluating the log materials provided by the participating teachers.

Initial efforts to construct a phenomena-based evaluation framework resulted in a multi-faceted matrix to be applied to each evidentiary phenomenon. This approach was intended to take into account both what was included in instruction and how students were engaged. We found that using this matrix to evaluate unit instructional logs led to unacceptably low inter-rater reliability, even after revising the approach to reduce the amount of inference required of raters. Using this approach also required a great deal of rater time.

Accordingly, we developed an even lower-inference approach in which a score is assigned to each identified phenomenon based on how it is addressed in the unit log. Any phenomenon addressed in a manner likely to promote student understanding is assigned a +1 score regardless of the frequency or depth with which it was treated. Any phenomenon addressed incorrectly is assigned a -1 score. Those phenomena not addressed at all are assigned a score of zero. The rating for a given unit log is the sum of scores assigned for all phenomena. Several HRI researchers and a content expert analyzed and rated the same three unit logs using this system. After reaching consensus, individual members of the rating team evaluated additional unit logs.

A total of 30 unit logs, a subset of the 107 collected, have been rated for the Plate Tectonics content area. Our analysis, using a three-level HLM model built with time-points nested within students nested within teachers, showed *no relationship* between these ratings and changes in student assessment scores.

We are currently analyzing instructional logs for the Force and Motion content area and anticipate presenting these findings at the conference in January.

B. Key insights (retrospective for veteran projects, prospective for newer projects) that have value for the Learning Network
   1. Our different findings for force and motion versus plate tectonics suggest the relationships among teacher knowledge, amount of instruction, and student learning depend on the content students are studying. Most surprising are the findings that (1) amount of instruction by itself does not predict student learning of force and motion concepts, and (2) neither teacher knowledge nor amount of instruction predicts student learning in plate tectonics.

2. This study also suggests that teacher instructional logs do not provide sufficient evidence to rate the quality of instruction and that direct observations of instruction may be necessary.

References

Druva, C. A., & Anderson, R. D. (1983). Science Teacher Characteristics by Teacher Behavior and by Student Outcome: A Meta-Analysis of Research. *Journal of Research in Science Teaching*, *20*(5), 467-79.

Goldhaber, D. D., & Brewer, D. J. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129-45.

Monk, D. H. (1994). Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement. *Economics of Education Review*, *13*(2), 125-45.

Wallace, M. R. (2009). Making Sense of the Links: Professional Development, Teacher Practices, and Student Achievement. *Teachers College Record*, *111*(2), 573-596.