Session Number: 1

Abstract Name: Using HLM to Study the Relationship between Teacher and Student Content Knowledge

MSP Project: Washington University Life Sciences Teacher Institute: Education for a Global Community

Author(s): Vicki May, Carl Hanssen, and Phyllis Balcerzak

Presenter(s): Vicki May, Carl Hanssen, and Phyllis Balcerzak

1. Questions(s) or issue(s) for dialogue at Learning Network Conference session:

What changes in teacher content knowledge have been demonstrated by Institute participants? Can these content knowledge changes be attributed to participation in the Institute? Have they been sustained?

What relationships can be detected between teacher content knowledge and student content knowledge? What student characteristics and/or classroom characteristics might help explain differences in student content knowledge?

2. Context of the work within the STEM education literature and within your MSP project:

An integral element of the Life Sciences for a Global Community (LSGC) Institute is a comprehensive assessment program targeted at participating teachers and their students. The goal of this program is to document teacher content knowledge changes as they progress through the Institute. Student content knowledge is also assessed to examine the relationship between teacher and student content knowledge. The underlying rationale for this work is that a direct relationship exists between teacher and student content knowledge. Thus, developing teacher content knowledge is a worthwhile strategy for improving student performance in science.

The STEM educational research supports a connection between high quality teacher professional development and teacher learning (Borko, 2004).

The LSGC Institute was designed to incorporate findings from the science education research. For example, most effective professional development programs for science teachers involve sustained inquiry over time. (Supovitz & Turner, 2000; Banilowe, Boyd, Pasley, & Weiss, 2005), strengthen science content and contextualizing the activities around the needs of the teachers in their classrooms (Garet et al., 2001; Cohen & Hill 2001), and provide opportunities for teachers to engage in scientific research processes (Silverstein et al., 2009). The design and decisions about implementation of the Institute were guided by these design principles.

Teachers first encounter the Institute through a three week on-campus residential program, with each week focused on a different content area that aligned with both the high school curriculum and the research of the lead scientist who organized the week’s instruction. Each course involved direct instruction by the lead scientist, who provided content current to the field and background
information about the research methods used to build knowledge in the field. Guided by the scientists, graduate students and post-doctoral fellows designed laboratory and field activities for teachers to experience the research models and reinforce subsequent content knowledge from direct instruction. Additionally, science educators designed laboratory investigations that connected the new knowledge to high school classroom instruction. The success of this model was dependent on, 1) the scientists embedding general knowledge of interest to high school teachers in the story of their own research, 2) the enthusiasm and passion of the graduate students/post-docs in their desire to have others understand what they know and do as young scientists, 3) the capacity of the science educators to translate key components to the teachers and scientists, 4) the teachers’ willingness to share classroom activities that support Institute content knowledge, and 5) the evaluator’s feedback to the project leadership team when links between in the pedagogical flow were not present or evident to teachers.

Similar to STEM data nationally, formative assessment of the fidelity of the Institute program implementation and summative data on teacher content learning before and after Institute coursework show positive impact across multiple variables. There is less empirical evidence in the published literature supporting the claim that teacher learning impacts student learning in K-12 classrooms. However, a recent study shows a positive improvement in the science achievement of students of teachers who participated in scientific research programs compared to the students of teachers who had not had these bench science research experiences (Silverstein, et al, 2009). The assessment program of the LSGC was designed with the intention of adding to the literature base that connects teacher learning to student learning in high school life science classrooms. This presentation is a reporting of findings in progress.

3. Claim(s) or hypothesis(es) examined in the work (anticipating that veteran projects will have claims, newer projects will have hypotheses):

The fundamental claim that is being explored is that teacher content knowledge is positively related to student content knowledge. This is the underlying premise for the argument that a critical strategy for improving student performance it to increase teacher content knowledge.

4. Evaluation and/or research design, data collection and analysis:

The LSGC has implemented an extensive assessment program for Institute participants (teachers) and their students. For teachers, two assessments were developed that are keyed to the content delivered during each summer institute (Summer I and Summer II). In year 1, each teacher took the Summer I pre-test and post-test. In year 2, teachers took the Summer II pre-test and post-test. In addition, they also took the Summer I post-test as a measure of knowledge retention. This strategy has been repeated for each teacher cohort. This staggered approach allowed the use later teacher cohorts as a comparison group for earlier cohorts.

Student tests are also keyed to summer institute content. Teachers are asked to administer these assessments to their students in fall and spring of each school year. The student test administered is keyed to the teacher’s progress through the institute. For example, in 2008-09,
teachers from Cohort 1 administered the Summer II student test because they had just completed the Summer II Institute Content. Teachers in Cohort 2 and 3 administered the Summer I content test; Cohort 2 completed the Summer I content while Cohort 3 teachers served as a comparison group.

This combination of teacher assessments aligned to student assessments, by content area provides an analytical opportunity to see overall effects of the teacher learning in the Institute on gains in student learning. The items for each assessment were selected, written and compiled by a team consisting of the project director, the lead scientist for each content area, and a statistician/educational psychologist. Each year the teacher and student assessments were submitted to validity/reliability testing with standard psychometric software.

The first year assessment contained more items than subsequent years so the most effective items could be used for subsequent testing. Initial modifications to the tests were made with the advice of the external evaluation team after item analysis. During years 2-3, the test items were submitted to psychometric testing each year, with minor changes in item wording and distracters. As advised by the project’s external evaluator, all changes were recorded on an item blueprint to inform decisions during analysis or interpretation of results. Each subsequent year resulted in fewer changes in items, while no changes were made to 30% of the items on each test so that longitudinal analyses and interpretations could be made with identical questions from year 1 to year 5. This process resulted in content tests that were flexible enough to assess learning as teaching goals change from year to year; and stable enough to be used in a longitudinal analysis.

This design enables a range of analytical options. First, simple pre-test post-test comparisons were made, indicating that teachers did demonstrate content knowledge gains between the beginning and end of the summer institute session. Similarly, the retention test was used to determine if those content knowledge gains were retained during the year after attending the Institute. As the assessment program continues, these analyses will be repeated.

Teacher Test Results

Figures 1 and 2 below summarize the teacher test results for the Summer I and Summer II teacher tests, respectively. For both teacher tests, LSGC participants demonstrated content knowledge gains after participating in the Institute.
Figure 1. Summer I Teacher Test Results

Figure 2. Summer II Teacher Test Results
As indicated in Figure 1, teachers from all Cohorts demonstrated knowledge gains from the pre to post-test for the Summer I content. However, teachers from Cohorts 1 and 2 did not demonstrate knowledge retention—their scores reverted back to pre-test levels. One possible explanation for this is that teachers were not necessarily utilizing the content knowledge gained during the Institute, i.e., Institute content may not be fully aligned with required school curricula (a tall order given that districts across the nation are served by the Institute).

As indicated in Figure 2, teachers from Cohorts 1 and 2 who took the Summer II test also demonstrated content knowledge gains from the pre-test to the post-test. Importantly, pre-test scores for those cohorts were not different than scores from comparison group teachers in Cohorts 2 and 3, who also took the Summer II test when they began the Institute, one year prior to attending the second summer institute. This strengthens the argument that the Institute is having a positive impact on teacher content knowledge in the immediate term. Utilization and reinforcement of that knowledge will likely contribute to knowledge retention.

**Student Test Results**

Figure 3 displays test results from students of teachers in Cohorts 1, 2, and 3. Students from all Institute teachers may take the test even though the teacher might not have begun attending the summer sessions. So, for test results from 2007-08, only Cohort 1 teachers had attended the institute, while Cohort 2 and 3 teachers were awaiting the start of their program. Thus, students of teachers in these cohorts serve as reasonable comparison groups to students of Cohort 1...
teachers. At the same time, students take the content test that is aligned with the summer institute content taught to teachers during the preceding summer. So, students of Cohort 1

*Figure 3. Student Test Results*

As shown in Figure 3, there is a consistent pattern of improvement for all students who took a content test. Students taking a content test in Spring scored higher on average than students taking a test in the Fall. Students taking both a Fall and Spring test also demonstrated gains.

*HLM Results*

Given the teacher and student results above, the next important question is to understand the proportion of variability in student test scores that could be attributed to teacher differences. HLM was used to isolate the variability in student content knowledge due to teacher (i.e., classroom) level variables. The primary dependent variable for these analyses was *spring student content knowledge*. The predictors at the classroom level (Level 2) included teacher content knowledge. Additional analysis are planned that will incorporate classroom characteristics as Level 2 predictors.
The first HLM model that was fit to the data utilized student spring test scores as the dependent variable. No teacher level predictors were included in the model, resulting in fitting an unconditional model, to determine the total amount of between student level variability that could be predicted by the inclusion of teacher level variables. The results of fitting this model are depicted in Table 1. The variance components in this table can be utilized to determine the intraclass correlation (ICC), which represents the proportion of variability in student level scores that exists between teachers, and therefore can be explained by including teacher level variables in level 2 of the HLM model. These results indicated that approximately 45 percent (ICC = 135/135 + 165) of the variability in student test score gains can be attributed to teacher level variables.

Table 1: Results from fitting the Unconditional Model to Student Test Scores

<table>
<thead>
<tr>
<th>Estimated Fixed Effects</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student test score (intercept)</td>
<td>43.8</td>
<td>1.67</td>
<td>26.12</td>
<td>48</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated Random Effects</th>
<th>SD</th>
<th>Var</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean student score by teacher</td>
<td>11.6</td>
<td>134.8</td>
<td>48</td>
<td>1740.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Student test score</td>
<td>12.8</td>
<td>164.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Once it was determined how much variability was due to between teacher differences, a conditional model was fit using teacher post-test score as a teacher level (i.e. Level 2) predictor to help explain differences in student spring test scores. This was based on the fundamental premise that teacher content knowledge contributes to student performance. The results of this analysis are shown in Table 2. As the table indicates, teachers that had higher post-test scores were more likely to have students with higher spring test scores. \( t = 2.23, p = .03 \). This indicates that including this variable helps to explain some of the between teacher variability. The variance components from the two models fit can be used to calculate an ICC which reflects how much variability is explained by including this Level 2 predictor in the model. Specifically, \( ICC = (135-117)/135 = .13 \). Therefore, approximately 13% of the between teacher variability can be explained by differences in teacher post-test score.

Table 2. Conditional Model Using Teacher Post-Test Score as a Predictor

<table>
<thead>
<tr>
<th>Estimated Fixed Effects</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student test score (intercept)</td>
<td>43.8</td>
<td>1.5</td>
<td>28.2</td>
<td>47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Teacher Post-Test Score</td>
<td>38.6</td>
<td>17.3</td>
<td>2.23</td>
<td>47</td>
<td>.030</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated Random Effects</th>
<th>SD</th>
<th>Var</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean test score by teacher</td>
<td>10.8</td>
<td>117.4</td>
<td>47</td>
<td>1481.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------</td>
<td>-------</td>
<td>----</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Student test score</td>
<td>12.8</td>
<td>164.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results provide initial evidence that the LSGC Institute is positively impact student performance—teacher content knowledge is being impacted by participation in the Institute, and teacher content knowledge helps to explain approximately 13% of the variability in student test scores.

Additional analyses are planned with the intent of explaining a greater proportion of the variability in student performance. First, as additional data become available through the assessment program, the analysis will be updated. For example, students of teachers in Cohort 3 will be incorporated into the analysis after they complete the Summer I content tests during the 2009-10 school year.

Second, the question of test alignment must be addressed in more detail. The current content tests incorporate six domains—Ecology, Evolution, and Plants & People for Summer I, and Matter & Energy, Genetics, and Neuroscience for Summer II. The test scores currently being used for analysis incorporate all three domains for each test, but high school curricula rarely address the Plants and Neuroscience domains. Thus, if the test scores are partitioned into sub-scales, a more precise explanation of the variability in student scores might be obtained.

Third, there are contextual factors that contribute to student performance that are not accounted for in this model. For example, socio-economic status of students is known to be related to student performance. Thus, if the database were portioned to examine just students who, for example, do not receive free-reduced lunch benefits, then a more complete story of Institute impact can be obtained.

Finally, teacher pedagogical performance, which is an important Institute goal, is not addressed in this work to date. It is reasonable to assert that teacher performance, as well as content knowledge, contribute to student performance. The introduction of teacher pedagogical metrics, then, will be important for refining this work.
5. Key insights (retrospective for veteran projects, prospective for newer projects) that have value for the Learning Network:

An important consideration for MSP projects is establishing a comprehensive assessment strategy to evaluate impacts on the direct target population as well as indirect targets. In the LSGC case, teachers are the direct and immediate targets. The assessment strategy being implemented allows us to assess content knowledge gains before and after Institute participation as well as retention of that knowledge after completing the Institute.

Students are indirect targets but Institute goals clearly articulate that gains in student content knowledge and improved attitudes towards science are desired outcomes. The path to these results, though, is through teachers who have been directly impacted by the Institute. Thus, the ability to closely link teacher and student assessments, and to apply appropriate methodologies for understanding those links, is important for evaluating Institute impact.

References


