

Comparing Teachers' Adaptations of an Inquiry-Oriented Curriculum Unit  
with Student Learning

Jay Fogelman and Katherine L. McNeill

University of Michigan

contact info:

Center for Highly Interactive Computing in Education

610 E. University Ave., Ann Arbor, MI, 48109-1259

734-647-4226

Paper presented at the annual meeting of the American Educational Research Association, April, 2005, Montreal, Canada.

This research was conducted as part of the Investigating and Questioning our World through Science and Technology (IQWST) project and the Center for Curriculum Materials in Science (CCMS), supported in part by the National Science Foundation grants ESI 0101780 and ESI 0227557 respectively. Any opinions expressed in this work are those of the authors and do not necessarily represent either those of the funding agency or the University of Michigan.

## Introduction

The science learning goals specified in national standards documents (American Association for the Advancement of Science (AAAS), 1993; National Research Council (NRC), 1996) have provided an opportunity for researchers to focus their efforts to develop classroom resources that enhance student learning on key learning goals. In addition to establishing a coherent framework for the science topics at the different grade levels, these documents suggest that students should learn science by engaging in inquiry processes that allow them an active role in their own learning and reflect how knowledge is constructed within the various scientific communities.

Recent reviews of traditional textbooks have called into question the degree that these textbooks support students developing deep understandings of the learning goals identified in the national standards (Kesidou & Roseman, 2002). To provide more effective classroom materials, researchers at the Center for Highly Interactive Curriculum, Computing, and Classrooms at the University of Michigan (hice) and the Learning Sciences Group at Northwestern University developed the Investigating and Questioning the World Through Science and Technology (IQWST) curriculum units (Reiser et al., 2003). One of the first two units designed for IQWST is a middle school chemistry unit, "How Can I Make New Stuff from Old Stuff?".

Early enactments of the "Stuff" unit in urban, suburban, and rural settings indicated that they helped teachers address their target learning goals successfully and supported student learning (McNeill et al., 2003). During these enactments, we observed teachers choosing to enact the unit's activities in different ways. This process of teacher adaptation, or transformation, is a common occurrence when teachers use innovative materials (Pinto, 2005), and an essential step if the materials are to be used long term in these classrooms (Fullan, 1982; Blumenfeld et al., 2000). We believe that it is important to understand more about how these adaptations of the unit's activities affect student learning. In this study, we used a post-enactment survey to ask teachers about their enactments of the "Stuff" unit and compared these results to what their students learned during the unit, as indicated by their test gain scores.

## Theoretical Framework

### *Role of Curriculum in Educational Reform*

Over the last decade, researchers have worked to incorporate what we currently know about teaching and learning into a definition of effective curriculum materials. Effective curriculum materials must meet the follow criteria: (1) Their content primarily focuses on a coherent set of important, age-appropriate student learning goals; (2) their instructional design effectively supports the attainment of the specified student learning goals; and (3) the teacher's guides support teachers in helping students attain these goals (Kesidou & Roseman, 2002). The first requirement reflects the need to focus on content specified either in the National Science Education Standards (NRC, 1996) or the Benchmarks for Science Literacy (AAAS, 1993). The second requirement specifies that the curriculum must call for instructional strategies that are

consistent with what we know about how people learn, such as calling for students to make sense of new experiences in light of what they already know, to share and refine their understandings, and assume responsibility for their own learning (Bransford et al., 1999). The third requirement is that resources be provided for the teacher so that he or she can facilitate an effective learning environment and develop expertise in areas such as students' commonly held ideas and strategies that could be used to assess student understanding. Because of the deficiencies of the texts used in most classrooms, there is a dire need for more effective science curricula (Kesidou & Roseman, 2002).

### ***Factors Influencing How Teachers Utilize Classroom Innovations***

In addition to trying to develop innovations that support student learning, researchers have recognized the importance of how teachers use these innovations with their students, and have turned their attention to teachers' adaptations, or transformations, of these materials. Analysis of past reform efforts indicated that in order for innovations to be sustained, teachers had to adapt them to meet local needs and conditions. In their review of past studies of teachers' adaptations of innovations, Pinto (2005) found that teachers' ability to adapt innovations was influenced by their knowledge and beliefs about the subject they were teaching, their beliefs about their own identity and about teaching and learning, and the degree that the innovation was supported within their local contexts. These factors are especially acute in impoverished schools, where there exists a "pedagogy of poverty" (Haberman, 1991). Innovations aimed at supporting classroom inquiry either through curriculum materials or technology are difficult to implement in such settings, due to inadequate space, materials, inadequate preparation and reflection time; low levels of science and computer knowledge, lack of training opportunities; large class sizes; high levels of teacher and student mobility; limited instructional freedom, lack of administration support, and unreliable internet connectivity (Songer et al., 2002).

Implementing classroom innovations requires teachers to change their practice and take on the unpleasant role of "novice" again (Fullan, 1982). A strong predictor of whether teachers can successfully meet this challenge is their sense of self-efficacy. Tschanen-Moran, Hoy, & Hoy defined a teacher's self-efficacy as his or her belief in their ability to act in ways that successfully accomplish specific teaching goals. In their review of the use of teacher efficacy, Tschanen-Moran et al (1998) found that teacher efficacy has been correlated with teachers' willingness to implement innovations.

Another factor that influences how innovations are enacted in classrooms is teachers' experience with the innovation. In past studies, we have seen that teachers continue to strengthen their use of reform-based curriculum materials through their second and third years using the units (Gier, 2005). Each time a teacher uses a particular innovation, we would expect an increase in both their understanding about how to use the innovation in his or her class, as well the effectiveness of their use of the innovative materials.

### ***Teachers' Curricular Adaptations***

The adaptations teachers make to innovations often diminish their intended function. Pinto (2005) identified common themes in the results from concurrent implementation studies of four

classroom innovations. Each team of researchers saw their innovations being transformed by teachers. Sometimes these transformations were benign and sometimes problematic. In all cases, the teachers tended to demote the goals of the innovation designers and adapt the innovation so its use more closely resembled familiar classroom practices. In order to provide opportunities for teachers to reflect on and refine their uses of innovations in subsequent professional development workshops, it is important to be able to share with them how specific transformations affect student learning. The transformations we are concerned with in this study include how much time teachers spend on the unit, the level of completion of the unit's activities, and whether they had students do the unit's investigations or present them as whole-class demonstrations.

When implementing a new curriculum or other classroom innovation, the teacher must decide how much time can be spent on the new unit. There has been considerable research on how time is spent in classrooms, and the effect of these practices on student learning. When using curriculum units designed to facilitate deep conceptual understanding, students need sufficient instructional time, i.e. time spent actively engaged in learning activities, to integrate their understandings. Reducing the amount of instructional time originally called for by the unit also reduces students' depth of understanding (Clark & Linn, 2003). Previous research on the effects of the amount of time that teachers allocate for particular classroom activities on student learning has produced mixed results, however. Since allocated time is not always spent on learning activities, these studies suggested that while allocating more time for particular activities may have a small positive effect for low ability students, there is no overall effect on what students learn (Cotton, 1989).

Teachers have to continually seek a balance between "covering" the topics they feel are important and ensuring that students' experiences are sufficient in order to develop deep understanding (Van den Akker, 1998). Teachers sometimes scale back student investigations, or decide to omit particular activities or portions of activities in the unit. Adaptations such as this might limit students' opportunities to engage in inquiry practices, such as asking questions and talking with classmates to solve problems, which result in greater student learning (Kahle, Meece, & Scantlebury, 2000). Consequently, we are interested in how the level of completion of a unit by a teacher influences student learning.

The tendency for teachers to transform innovative curriculum toward more traditional classroom practices suggests that how teachers choose to enact the unit might affect what students learn. The different ways that teachers manage classroom discourse have been called participation structures (Cazden, 1986) or activity structures (Fuson & Smith, 1998). These patterns of classroom discourse can vary in time scale and purpose, ranging from simple routines such as "initiation-reply-evaluation" (I-R-E) (Mehan, 1978,1979) exchanges where students answer questions and receive immediate feedback to a sequence of project milestones used to facilitate open-ended classroom inquiry (Polman, 2004). The tendency toward transmissive classroom routines despite accepted evidence for the need for students to take a more active role in their learning is well known (Bean, 2000). In other words, whole class teacher centered instruction often dominates classroom practice. We are interested in the relationship between teacher adaptations of the activity structures, such as whole class versus student centered, and student learning.

When teachers try to implement innovations such as standards-based curriculum units, there are many challenges. Teacher support structures are necessary for teachers as they as they implement reforms and refine their understandings (Fullan, 1982). Our own efforts at supporting systemic reform acknowledge and support teachers adapting innovative curriculum materials as they address the needs of their students, time constraints, and limitations in resources (Blumenfeld et al, 2000). One way that designers can support the adaptation process is by providing teachers with feedback on the effect their adaptations have, and to provide opportunities in subsequent workshops to reflect upon their practice and discuss enactment issues with colleagues and designers (Pinto, 2005). In this study, we ask the following research question: How do teachers' adaptations and understanding, specifically the amount of time on the unit, the level of completion of the unit, the activity structures, (i.e. whole-class demonstration versus student investigation), teacher self-efficacy, and whether a teacher has had experience enacting the unit before, influence student learning of target content and inquiry learning goals?

## Method

In order to address our research question, we used data from the enactment of the Stuff unit during the 2003-2004 school year. In this section, we begin by describing the Stuff unit in more detail. Then we discuss the participants and data sources that we used to address our research question. Finally, we describe our procedure for analyzing the test data and teacher survey data.

### *Description of “Stuff” Unit*

The IQWST curriculum units were designed to address the need for curriculum materials that support learning goals expressed in the national standards documents and to support classroom inquiry (Reiser et al., 2003). Aimed at middle school science classrooms, each IQWST unit includes a teacher's guide and student activity books that contain investigation sheets for each activity and reader passages that correspond to each lesson. The units' activities engage students in inquiry activities with relevant phenomena and support teachers in facilitating discussions that allow students opportunities to understand how their experiences relate to the units' learning goals. Each unit also includes supports for inquiry practices such as using evidence to construct scientific explanations and creating representations or models of phenomena.

The IQWST chemistry unit, "How Can I Make New Stuff from Old Stuff?" or "Stuff", introduces students to the concepts of characteristic properties, substances, chemical reactions, the conservation of mass during chemical reactions, as well as how the particulate nature of matter explains these macroscopic phenomena (McNeill, Harris, Heitzman, Lizotte, Sutherland, & Krajcik, 2004). The unit consists of 16 lessons, some of which contain several different activities. Some of the activities are identified as "optional," in order to provide teachers guidance in their adaptations of the completion of the unit. We felt that if teachers did need to cut activities in the unit because of time limitations that the optional activities could be removed and the students would still have some experience with all of the target learning goals. For example,

Lesson 13, “Does mass change in a chemical reaction?”, includes three activities. Activity 13A is an optional activity that has students investigate whether the mass changes when they create “gloop.” Activity 13.1 has students observing the reaction of Alka Seltzer in water in open and closed systems. Activity 13.2 has student redesign the 13.1 experiments so that mass will stay the same during the reaction. If all the “optional” activities are used, the unit is designed to take 33-35 days, but if only the “core” activities are used, the unit should take only 26-28 days.

### **Participants**

The 2003-2004 enactment of the “Stuff” unit included five different districts and 24 different teachers. We only included those teachers in the study from whom we received the required data sources, student pre and posttest data and the curriculum survey. This limited our analysis to 19 teachers (Table 1).

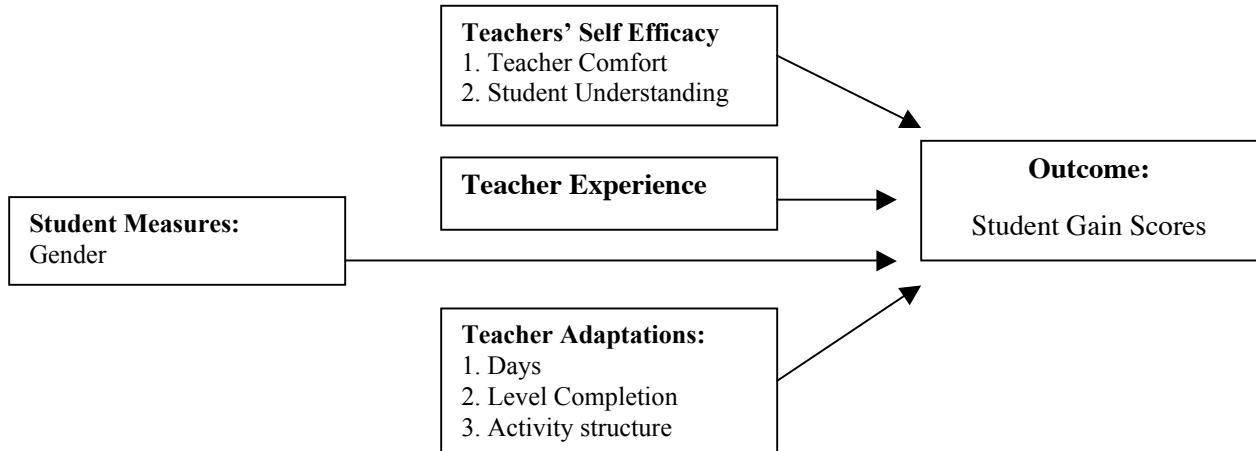
*Table 1: Participants from the 2003-2004 School Year*

Site	Urban A	Town B	Urban C	Suburb D	Rural E	Total
Schools	7	1	2	2	3	16
Teachers	8	3	2	3	3	19
Classrooms	30	5	4	13	13	65
Students	983	79	105	280	269	1716

Eight of the teachers were in public middle schools in a large urban area in the Midwest (Urban A). The majority of students in this school district is African American and come from lower to lower-middle income families. Three of the teachers taught in an independent school in a large college town in the Midwest (Town B). The majority of these students were Caucasian and from middle to upper-middle income families. Two of the teachers taught in a second large urban area in the Midwest (Urban C). The student population in this school district was 49.8% African American, 38% Hispanic, 8.8% Caucasian, and 3.2% Asian. Three of the teachers taught in a suburb of the second large urban area (Suburb D). The student population in this school district was ethnically diverse (approximately 42% Caucasian, 44% African American, 10% Hispanic and 4% Asian). Finally the last three teachers taught in a rural area in the south (Rural E). These schools had diverse populations each with a majority of African American students.

### **Measures**

To investigate the influence of teachers’ adaptations on students learning during the “Stuff” unit, we compared measures of student learning with factors that may influence teachers’ adaptations of the materials and the adaptation practices themselves. A conceptual model of our study including all of the measures that we investigated is shown in Figure 1. We describe each of these measures in more detail below.

*Figure 1: Conceptual Model*

**Description of Pre/Post Test.** To measure student learning for all teachers, the same test was administered to student before and after the “Stuff” unit. Only students who completed both the pretest and posttest were included in our analysis. Because of high absenteeism in the urban schools, only 1234 students completed both the pre and posttest. The test consisted of 15 multiple-choice items and 4 open-ended items for a total of 30 points. The test items were aligned with the learning performances of the unit, and therefore with the unit’s learning goals. All open-ended items were score using specific rubrics created to address the particular inquiry practice and content area. One rater scored all items. We then randomly sampled 20% of the tests, which were scored by a second independent rater. Our estimates of inter-rater reliability were calculated by percent agreements. Our inter-rater agreement was above 96% for each of the four open-ended test items.

In order to assess student learning over the unit, we used students’ gain scores. We calculated the gain scores by subtracting the pretest score from the posttest score. We used this measure as the outcome for our model. On the test, students also indicated their gender, which we also included in the model. Unfortunately, our agreement with the schools does not allow us to collect other demography data from the students so were not able to include race or other measures in our study.

**Description of Survey.** To gauge how teachers assessed and adapted the “Stuff” unit, each teacher was asked to complete a survey after they finished their enactment. The survey consisted of 16 pages, one for each of the unit’s lessons, which could include more than one activity (For a sample survey page, see Appendix A). Since we were interested in the teachers’ appraisals of their efficacy using the unit, they were asked to indicate their comfort-level with each activity and their students’ understanding of each activity. To get feedback on their adaptation strategies, teachers were asked to indicate whether each activity was done by students or as a teacher demonstration, its level of completion, and how many days were spent on each lesson. To determine each teacher’s experience with the unit, we used our records of previous enactments.

To analyze the survey responses, we first converted each teacher's checkmarks on the survey form to numerical codes and transferred them to a cumulative table. Table 2 summarizes how numbers were assigned to the teachers' responses.

Variable	Numerical Assignment
Self-Efficacy – Teacher Comfort Level	1 – low; 2 – medium; 3 - high
Self-Efficacy – Student Understanding	1 – low; 2 – medium; 3 - high
Experience	0 – first use of unit; 1 – second use of unit.
Teacher Adaptation – Activity structure	1 – teacher demo; 2 – student investigation; 3 - both
Teacher Adaptation – Level Completion	0 - not used; .5 - partially completed; 1 - completed
Teacher Adaptation - Days	Total number of days spent teaching the unit

Table 2: Numerical assignments for teacher's survey responses.

After tabulating teachers' responses, we reduced each teacher's responses to a single number for each of the variables listed above. For the teachers' self-efficacy, we averaged their responses for their own comfort level and their students' understanding for each activity across the entire unit. Each teacher's experience with the unit was coded as either their first or second use of the materials. In order to summarize the activity structures teachers used during the unit, we averaged their scores across all of the activities in the unit. For their level of activity completion, we totaled their scores across the unit and divided this total by the number of "core" or not optional activities. The total number of days each teacher allocated to the unit was found by adding the days he or she indicated were spent on each lesson.

### Analytic Method

Determining the impact of teacher adaptations on student learning is a complex issue, because each teacher's efforts only affect his or her students. In our analysis of the survey and test data, we needed to consider this grouping or nesting of students and any differential effects across teachers. Because of the multilevel nature of the data and our research questions, standard linear regression was insufficient for fully answering our questions. Regression requires independence of the units of analysis and ignores grouping leading to bias, incorrect estimation of standard errors, and underestimation of group effects. Multi-level modeling recognizes the dependence and grouping of data leading to more correct estimation of effects and variance. We used Hierarchical Linear Modeling (HLM) in a two-level format to investigate the effect of factors that affect teachers' adaptations and teachers' adaptation strategies on student learning (Raudenbush & Bryk, 2002). Our use of HLM consisted of three levels or steps. First, we created a fully unconditional model (FUM), then we created a level 1 or within-teacher model, and finally we created a level 2 or between-teacher model. Our level 1 variable is students and they are nested within our level 2 variable, teachers.

**Fully Unconditional Model.** HLM analysis begins with a fully unconditional model, which consists only of the outcome variable and no independent variables. In our model we used student gain scores, where we subtracted the pretest score from the posttest score, as the outcome variable. This model provided an estimate of the mean and confidence interval for the between teacher outcome measure ( $\gamma_{00}$ ). The fully unconditional model also provides the results of

partitioning the outcome variance into within-group ( $\sigma^2$ ) and between-group ( $\tau_{00}$ ) components, testing whether the between group component is significantly different from zero. If the between group proportion is significantly different from zero, this suggests that we should use an HLM analysis. In this case, we wanted to determine whether student learning is significantly different between teachers. From these measures we computed the intraclass correlation coefficient (ICC),  $\rho$ , which is the proportion of variation in the outcome measure that is due to differences between groups or between teachers.

**Within-Teacher Model.** After partitioning the variance and seeing if the between-teacher proportion was significantly different from zero, we began investigating the student-level measures that could account for the variation within teachers. Unfortunately, our agreement with the schools only allows us to collect students' gender and not other demographic data from individual students. Consequently, our level 1 model consisted only of gender. We entered gender as a fixed effect. This meant that the effect of gender did not vary depending on what teacher a student had. The following is the equation for our level-1 model:

$$\text{Gain Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Gender}_{ij} - \text{Gender..}) + r_{ij}$$

In this equation,  $\beta_{0j}$  represents the intercept or the gain score when all other variables equal zero,  $\beta_{1j}$  represents the effect of gender on student gain scores and, and  $r_{ij}$  represents the error term.

After running the within-teacher model, we can determine how much of the total unexplained individual-level variance for our outcome has been explained by the addition of our level-1 variable. This was computed by using the  $\sigma^2$  from the FUM we ran previously and the  $\sigma^2$  from this model.

**Between-Teacher Model.** Lastly, we ran a between-teacher model. This allowed us to model student learning with our teacher-level measures in an attempt to explain away the between-teacher variation in our outcome variable. More specifically, we could now determine if student learning differed by teacher adaptations. When running a between-teacher model, the level-1 model remains the same, but the level-2 model changes as we add teacher-level measures into the random-effect variable model to model the intercept. We tested the six teacher level variables that we described above: teacher comfort level, student understanding, whether a teacher had experience enacting the unit before, the number of days allocated to the unit, the level of completion of the unit's activities, and the teachers' activity structure, (i.e. whole-class demonstration versus student investigation). We removed any variables that were not significant. As a general rule, you need ten cases at a level (either level 1 or level 2) for each significant variable included in a model (Raudenbush & Bryk, 2002). Since we only had nineteen teachers in our study to include in the level 2 model, it was not surprising that we ended up with a model that included only two significant teacher practices. In our testing of the various models, we found two models that each included two significant variables. One model included teacher experience and level of completion and the second model included teacher experience and activity structure. Since the second model including teacher experience and activity structure had lower significant levels, we used it as our final model. Yet we hypothesize that if we had a larger sample of teachers, all three variables, teacher experience, level of completion and activity

structure, would significantly influence student learning. The following is our equation for the level-2 model for student gain scores:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Activity structure}) + \gamma_{02}(\text{Teacher Experience}) + \mu_{0j}$$

In this equation,  $\gamma_{00}$  represents the intercept,  $\gamma_{01}$  represents the effect of activity structure,  $\gamma_{02}$  represents the effect of teacher experience, and  $\mu_{0j}$  represents the error term. Teachers' activity structures (i.e. whole-class demonstration versus student investigation) and level of experience were used to model the intercept. None of the other teacher-level measures, days spent on the unit, student understanding, teacher comfort level or level of completion, were significant. Consequently, we removed them from our final level-2 models.

As with the within-teacher model, we can determine how much of the total unexplained individual-level and teacher-level variance of our outcome has been explained by the addition of our level-2 variables.

## Results

### *Descriptive Statistics*

Before creating our HLM model, we first examined whether there were differences in student learning and teacher adaptations across the 19 teachers. Table 3 displays the descriptive statistics for all of the variables included in our study.

*Table 3: Descriptive Statistics (n=1234)*

	Mean/% (Standard Deviation)
<b>Student Variables</b>	
Gender <sup>a</sup>	50.00
Test Gain Score	7.49 (5.23)
<b>Teacher Variables</b>	
Self-Efficacy – Teacher Comfort Level	2.55 (0.34)
Self-Efficacy – Student Understanding	2.39 (0.44)
Experience <sup>b</sup>	27.00
Teacher Adaptation - Days	31.17 (6.97)
Teacher Adaptation – Activity structure	1.93 (0.14)
Teacher Adaptation – Level Completion	0.94 (0.16)

a. Percentage of female compared to males

b. Percentage of teachers who have done the unit before compared to those who have not

Fifty percent of the students in the sample are male and fifty percent are female. We only included students in the analysis who completed both the pre and posttest. We computed each student's gain score by subtracting the pretest score from the posttest score. On average, students gained 7.49 points from the pre to posttest though the gain scores ranged from -13.36 to 22.80. For the teacher variables, we see a range of scores for both the teacher adaptation variables and the efficacy variables. For teachers' activity structure the average score was a 1.93. Remember

a score of 1 means that a teacher completed all activities as a demonstration, a score of 2 means that all activities were completed by students, and a score of 3 means that all activities were both demonstrated and students completed them. This suggests that for most lessons teachers had students complete the activities, but some were on average only completed as demonstrations. The average level of completion was 0.94 suggesting that typically teachers were completing a little less than the recommended core activities within the unit. On average, teachers spent 31.1.7 days on the unit. Twenty-seven percent of the teachers previously enacted the unit. Teacher's average comfort level was a 2.55, which is between medium and high. Finally, teachers' perception of student understanding was 2.39, which is also between medium and high.

Since we are interested in whether there is differential learning by teacher, we examined the effect size of student learning by teacher<sup>1</sup>. Figure 2 shows the effect sizes for the nineteen teachers.

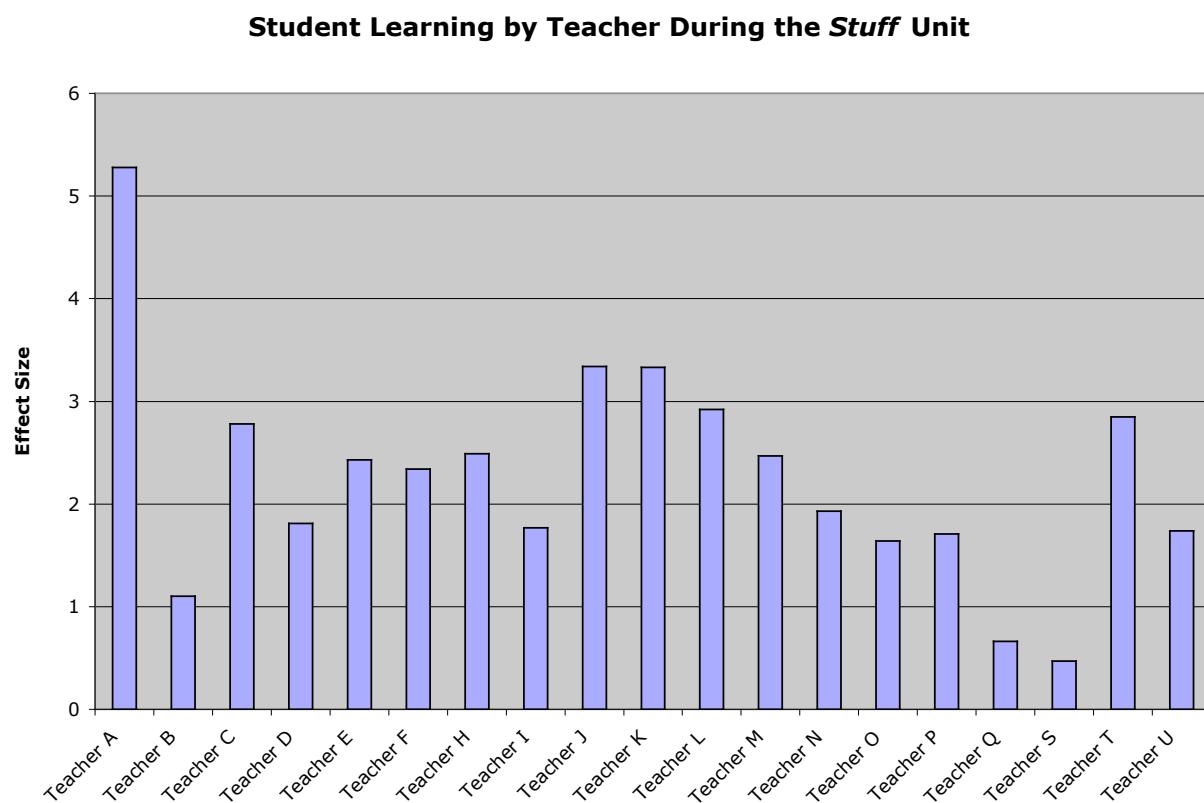


Figure 2: Effect size by Teacher

Across the nineteen teachers, there is a wide range of effect sizes from 0.47 to 5.27. We tested whether there was a significant teacher effect by performing an ANCOVA on students' posttest scores with the pretest scores as the covariate. There was a significant teacher effect with the learning gains of some teachers being greater than other teachers,  $F(18, 1215) = 9.062, p < .001$ . There was also a significant interaction between the teacher and students' pretest scores,  $F(18, 1215) = 2.868, p < .001$ , suggesting that the effect of a students' pretest on their posttest varied by teacher.

This analysis suggests that something is occurring in each of these classrooms that is influencing student learning. These differences could be caused by a variety of factors such as the school culture, parental influence, or different resources. We also believe that the differences in enactments are influencing student learning. Our hypothesis is that some of this difference in student learning is the result of teacher adaptations, experience and efficacy.

### **Fully Unconditional Model**

We began our HLM analysis by examining the fully unconditional model, which partitions the total variance in students' gain scores into its within- and between-teacher components. Table 4 provides the results from the unconditional model.

*Table 4: Unconditional Model of Student Learning  
(n=1234 students, N = 19 teachers)*

	Student Gain Scores
Tau ( $\tau$ )	0.384
Sigma-squared ( $\sigma^2$ )	0.646
Lambda-reliability ( $\lambda$ )	0.967
Intraclass Correlation (ICC) <sup>a</sup>	0.373
Adjusted-ICC <sup>b</sup>	0.380

<sup>a</sup>  $ICC = \tau / (\tau + \sigma^2)$   
<sup>b</sup> Adjusted ICC =  $\tau / (\tau + (\lambda\sigma^2))$

Lambda is the pooled reliability estimate across all the teachers for estimating our outcome variable, student gain scores. Since the reliability estimate is high, 0.967, we are comfortable using the adjusted intraclass correlation (ICC). The adjusted ICC tells us that 38% of the variance in student gain scores lies between teachers. There was a significant difference in student gains between teachers,  $\chi^2 = 693.85$ ,  $df = 18$ ,  $p < .001$ . Consequently, this supports our decision to use multilevel methods.

### **Within-Teacher HLM Model**

The within-teacher model explores whether gender is associated with student learning. We included gender as fixed effect, which means that the effect of gender did not vary depending on the teacher. Table 5 provides the results from our within-teacher model.

*Table 5: Within-Teacher Model of Student Gain Scores*  
(n=1234 students, N = 19 teachers)

	Student Gain Scores
<b>Random Effects</b>	
Intercept ( $\beta_0$ )	-0.012
<b>Fixed Effects</b>	
Gender <sup>a</sup>	0.104*
<b>Variance Components for Random Effects</b>	
Intercept variance ( $\beta_0$ )	0.383***

~ < .1; \* p < .05; \*\* p < .01; \*\*\* p < .001  
<sup>a</sup> Females compared to males

A student's gender does significantly influence their gain scores. On average, a female's gain score increases 0.104 standard deviations more than a male. Although adding gender does significantly influence student learning, it explains a very small percentage of the individual-level variance in student learning, less than 1%<sup>2</sup>. Unfortunately, we do not have access to other student level variables to include in the model. The intercept variance at the bottom of Table 5 suggests that there is still significant between-teacher variability. This provides support that there are contextual factors or characteristics of the teachers that influence student learning. In order to further unpack the role of teacher characteristics, we need to add level 2 predictors to our HLM model.

#### ***Between-Teacher HLM model***

Table 6 presents the results from our complete HLM model including both Level 1 and Level 2 predictors. Although we tested numerous teacher level characteristics in our model, we only kept in the model those measurements that were significant.

***Intercept as the outcome.*** The first set of results under intercept in Table 6 is for our model in terms of the intercept as the outcome. The reliability estimate for the intercept is 0.952 suggesting that the intercept is reliably estimated across all of the different teachers. These results tell us whether any of the teacher characteristics influence student learning.

*Table 6: Between-Teacher Model of Student Learning*  
(n=1234 students, N = 19 teachers)

	Student Gain Scores
<b>Random Effects</b>	
<b>Intercept (<math>\beta_0</math>)</b>	
Base	-0.234
Activity structure	1.869 <sup>~</sup>
Experience with Unit	0.715**
<b>Fixed Effects</b>	
Gender <sup>a</sup>	0.105*
<b>Variance Components for Random Effects</b>	
Intercept Variance ( $\beta_0$ )	0.258***

<sup>~</sup>< .1; \* p < .05; \*\* p < .01; \*\*\* p < .001

<sup>a</sup> Females compared to males

Teachers' activity structures (i.e. demo versus student investigation) has a marginally significant effect and a teacher's experience with the unit has a significant effect on student learning. Holding all other variables constant, as a teachers' activity structure increases by 1 point (i.e. goes from all lessons completed as demo to all lessons completed by students), students gain scores increase by 1.869 standard deviations. This is a very large increase in students' gain scores and suggests that having students actively complete the activities is important for students understanding of the key learning goals. On average, a teacher with experience teaching the unit has student gain scores of 0.715 standard deviations higher than a teacher who is completing the unit for the first time. This suggests the importance of having experience with reform based curriculum units.

Neither the number of days spent on the unit, teacher comfort level, teachers' report of their students understanding nor level of completion significantly influenced student learning. As we mentioned before, since our data includes only 19 teachers we would expect to only have at most two significant variables in our model. Our model is not powerful enough to detect significant effects of more variables. Other variables, particularly level of completion, which was significant by itself or in combination with teacher experience, could be important predictors of student learning if we had a more powerful model. Our model does not suggest that these other variables are not important; rather it provides support that both teacher experience and activity structure are particularly important for student learning.

**Proportion of between-level variance explained.** For average student learning between teachers, our model explains 33% of the variance.<sup>3</sup> By including only two variables in our model about teacher adaptations and experience, we explained a considerable percentage of the between teacher variation. Furthermore, we obtained the measure of teacher adaptation through a simple teacher survey of how they enacted the curriculum. Yet the variance component at the bottom of Table 6 shows that the between teacher variances is still significant. This means that we have not explained away all of the between teacher variance for student learning. There are other

measures not included in our model that explain why student learning varies with different teachers.

### **Conclusion and Implications**

This study used student pre/post tests and a teacher survey to investigate how teachers' adaptations, specifically the amount of time on the unit, the level of completion of the unit, the activity structures, (i.e. whole-class demonstration versus student investigation), teacher self-efficacy, and whether a teacher has had experience enacting the unit before, influence student learning of target content and inquiry learning goals. Student tests were used to determine student learning, and a multi-level analysis was performed on the survey data to determine the effect of teachers' curriculum adaptations on their students' learning.

We found the differences in student learning gains between teachers to be significant. Between-teacher variation accounted for 38% of the variance in student gain scores. Of the six teacher variables that we considered, only two of the variables, activity structure and teacher experience were found to be significant. Students who completed the activities had greater student gains than students in classrooms where the teacher completed the activities as demonstrations. Furthermore, experienced teachers had students with larger test gains. Our final model, which included the effects of activity structure and experience, explained 33% of the between-teacher variation in student learning gains. This result is consistent with previous studies documenting the importance of having students actively engage in making sense of their classroom experiences (Kahle, Meece, & Scantlebury, 2000) and the importance of teacher experience in enacting reform based curriculum (Gier, 2005). These results suggest that HLM analysis of survey data can contribute important knowledge about the impact of teacher adaptations of curriculum materials, which can be used by researchers as they bring their innovations to scale.

The relatively low number of teachers in our study limited the power of our model. We would expect the other adaptation measures, to be strong predictors of student learning, but their effects were not significant here. Specifically, we would expect that the level of completion of the unit and teacher efficacy measures to influence student learning. Our model does not suggest that these measures are not important, rather it just suggests that teacher experience and activity structure are particularly important. Though our model explained a considerable amount of the between-teacher variation, the remaining unexplained variance suggests that there are other important variables not in our model that explain why student learning varies between teachers.

Since our result accounted for a significant part of the between-teacher variation, this suggests that these results can be used to inform both teachers and researchers interested in studying curriculum adaptations.

One role of this study and surveys similar to this is to inform and support teacher practice. Supports such as educative curriculum (Ball & Cohen, 1996; Schneider et al, 2005) and sustained professional development opportunities help teachers refine their understandings of how to enact inquiry-based units (Blumenfeld et al, 2000). The relationships found between student achievement and teachers' accounts of their enactments can be used to bring research-based knowledge into teachers' conversations around subsequent enactments of the unit

(Fishman et al, 2003; Pinto, 2005). Specifically, teachers should understand the importance of activity structures and teacher experience. Spending more time on activities is not necessarily better, but that having students actually complete the investigations (compared to observing the activities as a demonstration) has a significant effect on what they learn. Teachers should also understand that student learning increases with teacher experience. This may be an important insight if teachers feel uncomfortable during their first enactment of the unit.

The second implication of this study is the possibility that teachers' adaptation strategies can be analyzed from afar by researchers interested in how innovations are adapted in classrooms. Pinto (2005) argues for the importance of systematic study of teachers' adaptations, and uses in-class observation methods to be able to describe in detail changes that teachers make to innovative materials. Though important, such an approach seems difficult to scale, limiting researchers' ability to study effects across a large number of teachers. Though more distal and limited in its ability to yield a rich account of teachers' specific adaptations, these results suggests that surveys such as this one used in conjunction with other methods like classroom observation, can help explain between-teacher variations in student achievement.

## References

- American Association for the Advancement of Science (1993). Benchmarks for science literacy. New York. Oxford University Press.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6-8, 14.
- Bean, J. A. (2000). Teaching in Middle Schools. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th Edition) (pp. 432-463). Washington: AERA.
- Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. w.(2000). Creating Usable Innovations in Systemic Reform: Scaling Up Technology-Embedded Project-Based Science in Urban Schools. *Educational Psychologist*, 33,3, (pp. 149-164).
- Bransford, J. D., Brown, A. L., Cocking, R. R. (1999). How people learn: brain, mind, experience, and school. Washington DC. National Academy Press.
- Cazden, C. B. (1986). Classroom Discourse. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd Edition) (pp. 432-463). New York: Macmillan.
- Clark, D., Linn, M. C.(2003). Designing for Knowledge Integration: The Impact of Instructional Time. *Journal of the Learning Sciences*, 12(4), (pp. 451-493).
- Cotton, K. (1989). School improvement research series: educational time factors. Retrieved April 1, 2005, from Northwest Regional Education Laboratory Web site:  
<http://www.nwrel.org/scpd/sirs/4/cu8.html>.
- Fishman, B., Marx, R., Best, S., Tal, R.(2003). Linking Teacher and Student Learning to Improve Professional Development in Systemic Reform. *Teaching and Teacher Education*.
- Fullan, M. (1982). The meaning of educational change. New York. Teachers College Press.
- Fuson, K. & Smith, S. (1998) The chalkboard activity structure as a facilitator of helping, understanding, discussing, and reflecting. Paper presented at the annual meeting of the American Educational Researcher Association, San Diego, CA.
- Geier, R. R. (2005). Student Achievement Outcomes in a Scaling Urban Standards-Based Science Reform. Unpublished doctoral dissertation, University of Michigan, Michigan.
- Haberman, M. (1991) The pedagogy of poverty versus good teaching. *Phi Delta Kappan*, 73, pp. 290-293.
- Kahle, J. B. Meece, J. & Scantlebury, K. (2000). Urban African-American middle school science students: Does standard-based teaching make a difference? *Journal of Research in Science*

Teaching, 37(9), (pp. 1019-1041.

Kesidou, S., Roseman, J.(2002). How Well Do Middle School Science Programs Measure Up: Findings From Project 2061's Curriculum Review. Journal of Research on Science Teaching, 39(6), (pp. 522-549).

McNeill, K. L., Harris, C. J., Heitzman, M., Lizotte, D. J., Sutherland, L. M., & Krajcik, J. (2004). How can I make new stuff from old stuff? In J. Krajcik & B. J. Reiser (Eds.), *IQWST: Investigating and questioning our world through science and technology*. Ann Arbor, MI: University of Michigan.

McNeill, K. L., Lizotte, D. J., Harris, C. J., Scott, L. A., Krajcik, J., Marx, R. W.(2003). Using Backward Design to Create Standards-Based Middle-School Inquiry-Oriented Chemistry Curriculum and Assessment Materials. Paper presented at the annual meeting of the National Association For Research in Science Teaching, 2003, Philadelphia, PA.

Mehan, H. (1978). Structuring school structure. Harvard Educational Review, 48, pp. 32-64.

Mehan, H. (1979). Learning lessons. Cambridge, MA: Harvard University Press.

National Research Council (1996). National science education standards. Washington. National Academy Press.

Pinto, R.(2005). Introducing Curriculum Innovations in Science: Identifying Teachers' Transformations and the Design of Related Teacher Education. Science Education, 89(1), (pp. 1-12).

Polman, J. L. (2004). Dialogic Activity Structures for Project-Based Learning Environments. Cognition and Instruction 22(4), pp. 431-466.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2<sup>nd</sup> Edition*. Newbury Park, CA: Sage Publications

Reiser, B., Krajcik, J., Moje, E., Marx, R.(2003). Design Strategies for Developing Science Instructional Materials. Paper presented at the annual meeting of the National Association of Researchers in Science Teaching, 2003, Philadelphia, PA.

Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005) Enacting reform-based science materials: the range of teacher enactments in reform classrooms. Journal of Research in Science Teaching, 42(3) pp. 283-312.

Songer, N. B., Lee, H., Kam, R. (2002). Technology-Rich Inquiry Science in Urban Classrooms: What Are the Barriers to Inquiry Pedagogy? . Journal of Research in Science Teaching, 39(2), (pp. 128-150).

Tschannen-Moran, M., Woolfolk Hoy, A. & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research* 68, 202 – 248.

Van Den Akker, J. (1998) The science curriculum: between ideals and outcomes. In B. J. Fraser and K. G. Tobin (eds.) *International Handbook of Science Education*. Great Britain: Kluwer. pp. 421-447.

## Appendix I: Sample Survey Page

### How Can I Make New Stuff From Old Stuff? Curriculum Questionnaire

Name: \_\_\_\_\_ School: \_\_\_\_\_ State: \_\_\_\_\_ Unit Start Date? \_\_\_\_\_ Unit End Date? \_\_\_\_\_

#### Learning Set I Properties and Substances: How is this stuff the same and different?

##### Lesson 1: How is this stuff the same or different? [Properties]

	Activities were done by		Level of completion was			Your comfort level was			Students' understanding was			Reader was used		
	teacher (demo)	students	completed	partially completed	not used	low	med	high	low	med	high	at home	in class	not used
1.1 Describing fat and soap														
1.2 Box of "Stuff"														
1.3 Initial concept map														

*Approximate number of days spent on this lesson:* \_\_\_\_\_ *In which of your classes did you use this activity?*

*Comments about lesson:*

*Describe any modifications to the lesson*

## Technical Notes

---

<sup>1</sup> ^Effect Size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation

<sup>2</sup> From the Fully Unconditional Model, we found that the amount of variance at the individual level was 0.64575. After taking into account our predictor variables in our within-teacher model, the within teacher variance is 0.64359. Therefore, the proportion of the individual-level variance that has been explained by our individual-level predictors is  $(0.64575 - 0.64359) / 0.64575$ , which equals 0.0033. This means that our within teacher model explains 0.33% of the variance in student learning.

<sup>3</sup> To calculate the proportion of the between-level variance that we explained in our model we used the following equation:  $(\tau_{\text{withinmodel}} - \tau_{\text{betweenmodel}}) / \tau_{\text{withinmodel}}$ . In this case  $(0.38257 - 0.25814) / 0.38257 = 0.3252$ .