The Linking Study—First Year Results

A Report of the First year Effects of an Experimental Study of the Impact of Feedback to Teachers on Teaching and Learning

Jonathan Supovitz
Graduate School of Education, University of Pennsylvania
Consortium for Policy Research in Education

The research reported in this paper was conducted by the Consortium for Policy Research in Education (CPRE). The opinions expressed in this paper are those of the author and do not necessarily reflect the views of CPRE, the subjects studied or the sponsors of this research study.

The Linking Study – First Year Results

A Report of the First year Effects of an Experimental Study of the Impact of Feedback to Teachers on Teaching and Learning

Jonathan Supovitz
Graduate School of Education, University of Pennsylvania
Consortium for Policy Research in Education
www.cpre.org

Paper Presented at AERA, Vancouver, BC
April 2012

Introduction

Districts expend a considerable amount of time, effort, and resources providing student performance data to teachers to inform instruction and examining those data for organizational decision-making. Yet rarely, if ever, are these outcome data connected to what goes on in classrooms. In this experimental study we report on the first year effects of an intervention designed to provide teachers with feedback on their teaching in conjunction with data on the learning of their students. The effects on subsequent teaching and student learning are examined in comparison to teachers who received feedback only on the learning of their students.

Background

The recent preoccupation with "data-driven decision making" in education is based upon both longstanding and contemporary trends. Policymakers have long had interest in using evidence to inform decision-making (Simon, 1955; March & Olsen, 1976). More recently, the accountability movement in general, and the No Child Left Behind (NCLB) Act of 2001 in particular, has increased demands for testing in both states and districts and the local use of those data for improvement (Elmore, Abelmann, & Fuhrman, 1996; Hamilton, Stecher, & Klein, 2002). In addition, the rapid proliferation of technology-based tools has made the collection, aggregation, and organization data increasingly available (Stringfield, Wayman, & Yakimowski-Srebnick, 2005).

The delivery of data to the classroom, however, does not guarantee its productive use. Several studies point to the need for instructional support to supplement insights from the data. Research that has looked more deeply at periodic assessment has found that information alone is not a solution to weak instruction. A 2005 RAND report on data use in Pennsylvania, for example, found that many teachers and principals who express interest in data for improving instruction "lack sufficient data analysis skills and a process for systematically using data." (Dembosky, Pane, Barney, & Christina, 2005, p. 47). CPRE's own study of interim assessment use in two districts has found that while a great amount of resources are dedicated to collection and scoring of data, much less is devoted to analysis of assessment results, and there is almost no systematic support for instructional modification (Oláh, Lawrence, & Riggan, 2010). This lack of attention to the linkage between data generation and data use is a crucial void to be filled. Teachers vary greatly in their ability to use data for instructional improvement (Armstrong &

Anthes, 2001; Fuchs, Fuchs, Hamlett, & Stecker, 1991; Heritage, Kim, Vendlinski, & Herman, 2008), yet it is precisely through instructional modification that we hope to see increased student learning. As a result, many are proposing that data-use plans include professional development for both analysis and instruction (Graney & Shinn, 2005; Tharp-Taylor, Nelson, Dembosky, & Gill, 2007).

About the Linking Study

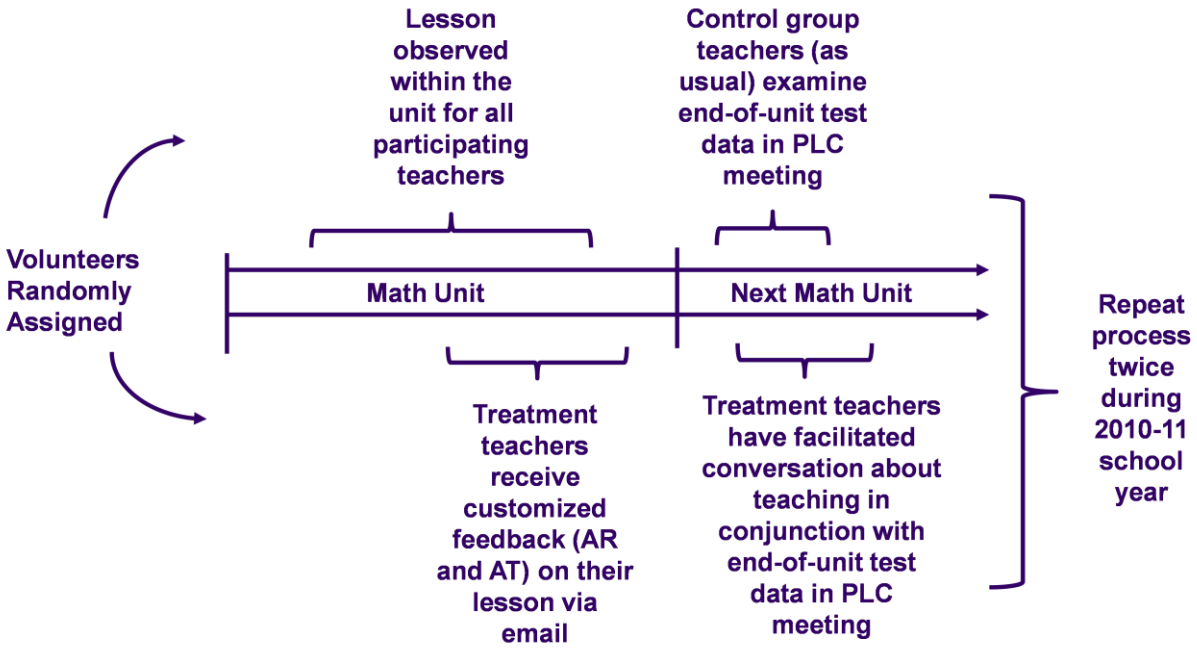
The Linking Study is funded by the Spencer Foundation of Chicago. The study is based upon the idea that teachers currently receive lots of data on the learning of their students (ie test scores) and are asked to make inferences back to their teaching. However, they receive little or no structured feedback on their teaching which produces the performance of their students. The hypothesis we are testing in the study is that feedback on instruction examined in conjunction with data on student learning is a more powerful condition to improve the quality of subsequent teaching than teacher examination of data on student learning alone.

In 2009, working with a medium-sized suburban school district, we designed an intervention to provide teachers data on their instruction, in the form of feedback on an observed lesson that they could examine in conjunction with end-of-unit test data. The district uses the Investigations mathematics curriculum in elementary schools, so that we focused our instructional feedback around two dimensions of mathematics instruction identified and measured by the Learning Research and Development Center (LRDC) at the University of Pittsburgh. The LRDC's measure, called the Instructional Quality Assessment, or IGA, measures two dimensions of instructional quality: Academic Rigor and Accountable Talk. We chose these dimensions because there was both research on their leverage to change instruction (Cobb, Boufi, McClain, & Whitenack, 1997; O'Connor & Michaels, 1996; Tharp & Gallimore, 1988) and because there were rubrics developed to assess them by the (Matsumura, Garnier, Pascal, & Valdés, 2002; Junker, Weisberg, Matsumura, Crosson, Wolf, Levison, 2005; Boston, & Wolf, 2006).

The goal of the project was to experimentally test hypothesis that data on teaching and learning was more powerful for teachers than data on learning alone. As the project proceeded through its first year, we found that the data collected from the study was providing us with lots of feedback that we could use to strengthen the intervention. We didn't hesitate to use this feedback to strengthen the intervention throughout the year, even as we continued the experiment.

To implement the intervention, teachers of mathematics in grades 3-5 were recruited in the Fall of 2010 to participate in the research project. Fifty teachers agreed to participate in the 2010-11 school year. The teachers represented 20 grade groups and, since feedback was to occur in PLCs, grade levels were randomly assigned to the treatment or control condition. An overview of the study design is shown in Figure 1. Regardless of their assigned group, all teachers had mathematics lessons videotaped in two pre-determined *Investigations* mathematics units, one in October 2010 and one in February 2011. The treatment group received both customized feedback on the Academic Rigor and Accountable Talk of their lessons via email, and a facilitated conversation within a subsequent professional learning community (PLC) meeting. During the PLC, they also examined their student end-of-unit test scores. The control group had a facilitated PLC conversation on their end-of-unit test scores, but no feedback on their instruction. A graphical representation of original study design looks something like this:

Figure 1. Overview of Linking Study Design



Data Sources

To measure the impacts of the intervention, we collected seven distinct sources of data. These are encapsulated in Table 1. The data sources for the study included surveys both before and after the project, short 'exit slip' surveys that were collected immediately following the professional learning community (PLC) meetings in which data were discussed, which occurred two times in the 2010-2011 school year. The data from the videotaped classroom observations conducted twice each year with each teacher, and both end of unit test data and end of year state test data. In addition, we also collected interview data from both the PLC facilitators and a sample of teachers.

Data for the study were collected in several ways. Table 1 summarizes the data collected over the course of the year. First, on-line surveys were administered to all 50 participating teachers in the Fall before they began participation in the project, and again in the Spring after participation. Second, after their professional learning community (PLC) group experience, both teachers in the treatment and control groups completed a short "exit slip" survey that focused on their reactions to their PLC experience. Third, the lesson that we videotaped in order to write up feedback were coded on multiple dimensions of academic rigor and accountable talk. Fourth, we conducted focus groups with the PLC facilitators in between the two rounds of feedback. Fifth, we conducted interviews with a sample of teachers in the treatment group to better understand their perceptions of the experience and their reports of the influence of the experience on their subsequent instruction. Finally, we collected two types of test data on all of the students of teachers who participated in the study. First, we collected the end-of-unit test data for each of the grade level units in 2010-11, regardless of whether it was the focus of the feedback session. Second, we collected the New Jersey state test data for all students whose teachers participated in the study.

Table 1. Summary of Data, Frequency, and Information Provided

Instrument	Frequency and Timing in 2010-11 year	Information Provided by Source
1. Online survey	Pre (09/10) – post (05/11)	Demographics of participants; attitudes and beliefs about data, unity of their grade-level PLCs, experience with *Investigations* math curriculum.
2. PLC exit slips	2x/year - Immediately following PLC meetings where teachers discussed data	Participant reactions to the PLC experience; perceptions of insight gained on teaching and student understanding; comfort examining data in groups.
3. classroom lesson observation	2x/year - During units selected for the study	Videotape of two lessons; data used to provide feedback to study participants as well as to rate quality of lesson in terms of Academic Rigor and Accountable Talk
4. Focus group with PLC facilitators	1x/year In between the two rounds of the intervention	Feedback on the facilitation process.
5. Interviews with treatment teachers; observations of PLC sessions.	3x/year – Before experience, after first round, after 2nd round.	Understanding of the experience and participant perceptions of the intervention and self reports on subsequent impacts.
6. End-of-unit assessment data	Collected online by the district at the end of each instructional unit.	Impacts on students.
7. NJ state test data	Available in the Fall of 2011	Another measure of impacts on students. Not provided in time to provide information at all.

Results

In this paper I report on data from three of the data sources in the table above. Additional project publications will focus on other aspects of the data collection. First, I discuss results from the exit slip survey that teachers completed immediately after their PLC meeting. Second, I present external raters' judgments of academic rigor and accountable talk. Third, I report preliminary results of student performance results.

1. Teacher Reported Perceptions of their Experiences in Professional Learning Communities

Using survey items from the professional learning community (PLC) exit slip survey, we examined teacher reports of impacts of their PLC experience. The exit slip survey was administered to all PLC participants immediately following their PLC meeting examining data. The PLC survey consisted of 20 items. Using a subset of these items we constructed three scales. The first scale, *Learning about Instruction*, was made up of five items that asked teachers about the extent to which they learned about their instruction in their PLC meeting examining data. The scale had a cronbach's alpha reliability of .89.

The second scale, *Learning about Students*, was a six item scale that asked teachers about the extent to which they learned about their students in their PLC meeting examining data. The alpha reliability of this scale was .78. The third scale, called *PLC Group Interaction*, consisted of four items which asked teachers about their comfort discussing data in their PLC and the quality of the conversation. This scale had an alpha of .72. All three scales had reasonable reliabilities above .70. The complete set of items for each scale is shown in Appendix A.

The pattern of results from the exit slips for teachers in the treatment and control groups are shown in Table 2. The results indicate that teachers in the treatment group did not feel they gained from the experience examining teaching and learning data more than did teachers in the control group that just examined their students' end of unit test data. In both the Fall and Spring cycles, there were no statistical differences in teacher reports about learning about their instruction in their PLC meeting examining data. In both cycles, teachers in the control group reported that they learned more about their students from examining data in their PLC meeting. This might not be surprising if we consider that the teachers in the control group spent the whole 40 minute PLC meeting examining their students' test data, while the teachers in the treatment group spent only a proportion of the meeting examining student test data.

Table 2. Participant Responses on Exist Slips about their Learning in PLC Meetings in which they examined data.

	Cycle 1 Fall 2010		Cycle 2 Spring 2011	
Instructional Dimension	Treatment (n=21)	Control (n=24)	Treatment (n=22)	Control (n=19)
Learning about Instruction (alpha = .78)	3.32 (.11)	3.34 (.49)	3.49 (.40)	3.57 (.32)
Learning about Students (alpha = .89)	3.56 (.51)	3.67 * (.45)	3.10 (.66)	3.61 * (.35)
PLC Group Interaction (alpha = .72)	3.27 (.55)	3.56 ~ (.47)	3.45 (.57)	3.67 (.44)

~ p < .10 * p < .05 ** p< .01

The final row from Table 2 shows that in both cycles the teachers in the treatment group reported that they were less comfortable examining data with their colleagues in the PLC meeting. In the Fall this difference was marginally significant (at the .10 level), while it was non-significant, but still favoring the control group) in the Spring cycle. This indicates that teachers in the treatment group, who were examining data on their instructional practice with their colleagues, felt *less* positive about examining data in their PLC groups compared to the control group.

In sum, from examining the exit slip data we get a picture of very little impact of the treatment. Treatment teachers felt no differently about what they learned about their instruction. The control group teachers reported in both the fall and spring cycles that they learned more about their students than did the treatment teachers. These significant differences might be explained by the fact that a greater proportion of the control group's PLC time was spent on looking at student test data because they did not examine data on instructional practice in their PLC. Finally, in both the Fall and the spring, the treatment group teachers reported less comfort examining data in their PLC than did the control group (although this difference was only marginally significant in the Fall and non-significant in the Spring). Together, these results show that, from the teachers' perspectives, the treatment did not teach

them more about their instruction or their students, nor did they feel comfortable doing these activities with their colleagues. As we will see, these results can be interpreted differently when looked at with the advantage of additional data.

2. External Lesson Ratings

Distinct from teacher perceptions of the value of their experiences examining data, we also have evidence from expert ratings of classroom observations. Lessons of both treatment and control group teachers were videotaped - one lesson in the fall of 2010 and one lesson from the spring of 2011 - in a designated curricular unit (different for each grade level, but common for all teachers within a grade level).

The focus of our assessments of instructional quality was based upon dimensions of the Instructional Quality Assessment (IQA), a rating of two dimensions of instructional quality, Academic Rigor and Accountable Talk. A rating system of Academic Rigor and Accountable talk was developed and validated by the Learning Research and Development Center at the University of Pittsburgh. Appendix B provides a brief summary of the dimensions of Academic Rigor and Accountable Talk. Our raters received one full day of training by a qualified IQA trainer who traveled to Philadelphia to conduct the training.

The data from the videotaped lessons was used for both part of the treatment and part of the research. The videotaped lessons were the basis for providing qualitative feedback to teachers in the treatment group about their instruction (see sequence of feedback in Figure 1). The lessons for all teachers were also rated by trained raters based upon three dimensions of Academic Rigor and four dimensions of Accountable Talk.

Table 3. External Ratings of Academic Rigor (AR) and Accountable Talk (AT)

	Cycle 1 Fall 2010		Cycle 2 Spring 2011	
Instructional Dimension	Treatment (n=24)	Control (n=24)	Treatment (n=21)	Control (n=22)
AR1: Potential of Task	3.75 (.12)	3.71 (.11)	3.80 ** (.12)	3.00 (.20)
AR2: Implementation of Task	2.96 (.15)	3.13 (.15)	3.40 * (.18)	2.81 (.19)
AR3: Student Discussion following the Task	2.50 (.26)	2.29 (.24)	3.00 * (.22)	2.43 (.29)
AT1: Student Participation	3.96 (.04)	3.58 (.10)	3.80 ~ (.12)	3.38 (.20)
AT2: Rigor of Teacher's Questions	3.29 (.18)	3.04 (.15)	3.45 * (.20)	2.75 (.20)
AT3: Teacher's Press for Knowledge or Thinking	2.92 (.19)	3.13 (.14)	3.35 ~ (.20)	2.75 (.25)
AT4: Student's Providing Knowledge or Thinking	2.96 (.20)	3.00 (.16)	3.20 (.23)	2.69 (.24)
AR Overall	3.07 (.12)	3.04 (.13)	3.40 * (.14)	2.75 (.17)
AT Overall	3.28 (.13)	3.19 (.11)	3.45 * (.14)	2.89 (.20)

~ p < .10 * p < .05 ** p< .01

The data in Table 3 show the mean and standard deviation of the external raters' assessments of teachers on the different dimensions of Academic Rigor and Accountable Talk. As you can see, there were no significant differences between the treatment and control group on any of the dimensions of Academic Rigor and Accountable Talk at the time of the first videotaped lesson in the fall. This is evidence that the random assignment process worked, because at the time the fall videotapes were taken, no treatment had occurred.

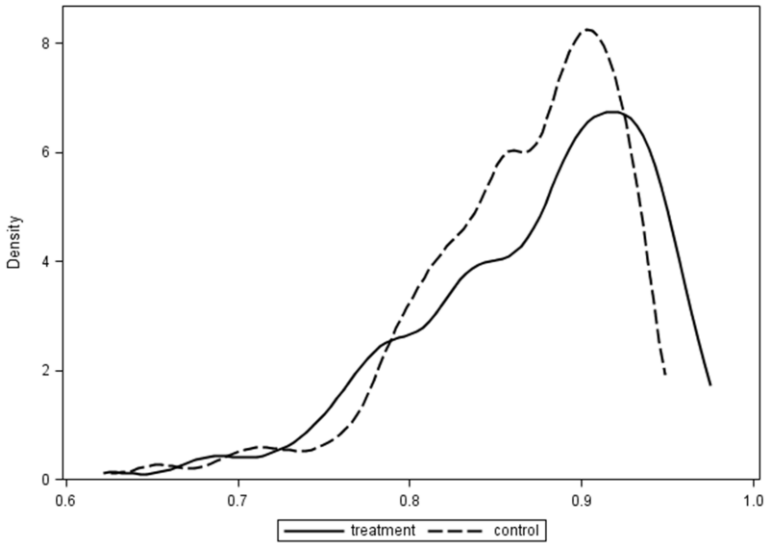
The lack of differences in the fall ratings are in striking contrast to the results from cycle 2 in the spring. The spring ratings, which were assessments of lessons that occurred after the first round of feedback but before the second round of feedback, show consistent and statistically significantly higher ratings for the teachers in the treatment group in comparison to the ratings of teachers in the control group. These differences are apparent in all three dimensions of Academic Rigor and all four dimensions of Accountable Talk, as well as in the cumulative measure of each dimension. These significant differences represent a widening of the gap between the treatment and control groups from cycle 1 to cycle 2. That is, there are some dimensions where the treatment group was roughly the same but the control group declined; other dimensions where the treatment group increased and the control group stayed static; and other dimensions where the treatment group increased and the control group decreased. These experimental results provide strong evidence of an impact on the quality of instruction associated with the feedback intervention.

3. Student Impact Results

Student achievement data in grades 3, 4, and 5 was analyzed for causal impact in a cluster randomized trial. In this experiment, two rounds of instructional training were given to teachers in the treatment group. The outcome in this analysis was derived by averaging student end-of-unit test scores following the two rounds of treatment. A measure of prior student achievement was calculated by averaging the scores from the units preceding the intervention. 798 students were present at some point during the school year, of which 775 (97.1%) were included in the analysis.

The impact of the treatment was estimated using multiple regression, in which prior student achievement and a teacher-level indicator of experimental condition were regressed on student outcomes in the unit following the treatment. The model also accounted for the multilevel nature of the data by including a random effect for teacher in order to partition the student outcome variation between and within teachers. One tenth (11.4%) of the total variation in student achievement was found to be related to the teacher, of which 47.1% was explained by the predictor

Figure 2. Distribution of achievement scores by treatment group, controlling for individual prior achievement and clustering of students within teacher.



variables. The results indicate that there were positive and significant main effects for both prior achievement (b=0.50, p<0.01) and treatment (b=0.022, p<0.05). Students in the experimental treatment group had significantly higher average unit scores following the treatment compared to the control group. These are represented in the graphic in Figure 2. The dashed line represents the distributed of the control group, while the solid line represents the distribution of the control group.

Discussion

Efforts to increase teacher data use in the United States over the past five years have been largely focused on increasing teachers' access to student test results. We hypothesize that while student test data are useful to teachers, they are a limited form of evidence because they provide no information for teachers about the teaching that produces student test results. In this study we consider an alternative perspective about what data might be useful to teachers by providing teachers with data on their teaching to examine in conjunction with data on the learning of their students.

Using a rigorous experimental design, this study examined the impacts of these different data conditions on teachers' perspectives of value, their instructional practice, and the learning of their students. The promising results reported in this paper present our first analysis of the impacts of the intervention that links data on teaching and learning for teachers. The results indicate that providing feedback to teachers on their teaching in conjunction with data on the learning of their students has more impact on subsequent teaching and student performance than providing data to teachers on student learning alone.

A provocative additional finding from these results is that teachers who were participating in the treatment did not perceive any additional value from the experience as compared to the teachers in the control group. On self-reported measures of learning about instruction and student understanding, we found that teachers receiving the treatment did not feel they gained additional knowledge or expertise. In fact, control teachers reported gaining more about their students' learning in their professional learning community (PLC) meetings than did the treatment group; a finding we attribute to their spending more time looking at student test data than did teachers in the treatment group. In addition, teachers in the treatment group reported feeling less comfortable examining data in PLCs than did teachers in the control group. We attribute this to teachers' discomfort with looking at data on instruction with their colleagues.

The external impact data we collected tell a very different story than did the teachers' self-report data. Comparing the external trained raters' judgments of the quality of instruction, assessed on scales of the academic rigor and accountable talk in videotaped lessons, we found that the treatment caused significantly higher levels of subsequent practice. Further, we found that the students of teachers in the treatment group had significantly higher levels of student performance than did the students of the control group teachers, after controlling for prior achievement.

These results suggest that efforts to provide teachers with richer data that explicitly links their instructional practices to the learning of their students is a promising way to expand the concept of teacher data use beyond a focus on test scores alone. We are currently replicating the experiment with a larger pool of teachers and a more refined and robust treatment to see if these results can be duplicated.

References

Armstrong, J., & Anthes, K. (2001). How data can help: Putting information to work to raise student achievement. *American School Board Journal, 188*(11), 38-41.

Boston, M., & Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of the Instructional Quality Assessment toolkit* (CSE Technical Report No. 672). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education, 28*(3), 258-277.

Dembosky, J. W., Pane, J. F., Barney, H., & Christina, R. (2005). *Data-driven decisionmaking in Southwestern Pennsylvania school districts.* WR-326-HE/GF. Santa Monica, CA: RAND.

Elmore, R. F., Abelmann, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and students achievement in mathematics operations. *American Educational Research Journal, 28,* 617-641.

Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review, 34,* 184-201.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education.* Santa Monica, CA: RAND.

Heritage, M., Kim, J., Vendlinski, T. P., & Herman, J. L. (2008). *From evidence to action: A seamless process in formative assessment?* Report 741. Los Angeles, CA: CRESST.

Junker, B., Weisberg, Y., Matsumura, L. C., Crosson, A., Wolf, M., & Levison, A., (2005). *Overview of the instructional quality assessment* (CSE Technical Report No. 671). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation.

March, J. G., & Olsen, J. P. (1975). The uncertainty of the past: Organizational learning under ambiguity. *European Journal of Political Research*, *3*, 147-171.

Matsumura, L. C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment, 8,* 207–229.

Oláh, L. N., Lawrence, N., & Riggan, M. (2010) Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education*, 85, 226–245.

O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussions. In D. Ghicks (Ed.), *Discourse, learning, and schooling* (pp. 63-103). New York: Cambridge University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99-118.

Stringfield, S., Wayman, J. C., & Yakimowski-Srebnick, M. E. (2005). Scaling up data use in classrooms, schools, and districts. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons learned from technology-based educational improvement* (pp. 133-152). San Francisco: Jossey-Bass.

Tharp, R. G., & Gallimore, R. (1988). Rousing minds to life: Teaching, learning, and schooling in social context. Cambridge: Cambridge University Press.

Tharp-Taylor, S., Nelson, C. A., Dembosky, J. W., & Gill, B. (2007). *Partners in Pittsburgh public schools' Excellence for All initiative: Findings from the first year of implementation.* Documented Briefing.

Appendix A. Exit Slip Scales and Contributing Items

LEARNING ABOUT STUDENTS SCALE* (ALPHA =.89)

1. The data we examined today gave me useful insights into the performance of my students.
2. I learned something today about the mathematics content of the unit we discussed.
3. The conversation in today's meeting helped my PLC get on the same page about mathematics instruction.
4. The data we examined on student performance gave me useful insights into the understanding of my students.
5. I gained a better understanding of how to examine student test data for insights into student thinking.
6. I plan to make changes in my teaching as a result of things I learned from examining student performance data

LEARNING ABOUT INSTRUCTION SCALE* (ALPHA =.78)

1. I learned something today about designing challenging math lessons.
2. I learned about engaging students to explain their thinking about how they solve mathematics problems.
3. I learned something today about developing students' conceptual understanding of mathematics.
4. I learned new strategies to press students to explain their thinking.
5. I plan to make changes in my teaching as a result of things I learned in this PLC meeting.

PLC GROUP INTERACTION SCALE* (ALPHA =.72)

1. The conversation in today's meeting helped my PLC get on the same page about mathematics instruction.
2. I would have preferred to examine these data on my own instead of with my grade level team. (REVALENCED)
3. Examining data with colleagues made the meeting more meaningful than examining the data on my own.
4. Please rate the overall quality of the discussion in your PLC today (3 point scale of Lo, Medium, Hi Quality)

*All responses on a four point scale (strongly disagree, disagree, agree, strongly agree) unless otherwise specified.

Appendix B. IQA Rubric for Mathematics

The IQA, developed at the University of Pittsburgh, Learning Research and Development Center, is a set of rubrics for looking at classroom instruction. During the 2010-2011 school year, 3rd-, 4th-, and 5th-grades across the district will use it as a *learning tool,* with the goal of better supporting mathematics instruction.

The two main areas of the IQA (and six sub-areas), along with their definitions, are given below. Each definition is followed by *one* example of excellent instruction. Of course, there are many ways to assist students in developing mathematical understanding. The goal of the IQA is to provide one structure for talking about quality mathematics instruction.

Academic Rigor

Potential of Task
Did the task have the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships? • *A good example would be a task of high cognitive demand (such as an Investigations task) that does not explicitly suggest the use of a specific procedure or strategy. The students should be engaged in "doing mathematics;" not simple recall or applying a procedure. The task should also require that students explain and/or record their thinking.*
Implementation of the Task
At what level did the teacher guide students to engage with the task in implementation? • *A good example would involve the students engaging in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships. The students should be 'doing mathematics' or applying procedures that are closely connected to mathematical concepts.*
Student Discussion Following the Task
To what extent did students show their work and explain their thinking about the important mathematical content? • *A good example would involve students showing/discussing more than one strategy or representation for solving the task and providing explanations of why the different strategies/representations were used to solve the task.*

Accountable Talk

Participation
Was there widespread participation (i.e., a response to a mathematical question) in teacher-facilitated discussion? • *Over 75% of the students participated throughout the discussion.*
Questioning
Does the teacher ask academically relevant questions that provide opportunities for students to elaborate and explain their mathematical thinking? • *For example, the teacher asks students to identify and describe the important mathematical ideas in the lesson.*
Asking (Teacher Press)
Were students pressed to support their contributions with evidence and/or reasoning? • *For example, the teacher almost always asks students to provide evidence for their contributions or to explain their reasoning.*
Providing (Student Responses)
Did students support their contributions with evidence and/or reasoning? • *For example, students consistently explain their thinking using reasoning in ways appropriate to the discipline.*