MARCH 2007

Research Brief III

Improving Instruction:

Developing the Knowledge and Skills of School Leaders

Center for Educational Leadership,
Norwalk-La Mirada Unified School District
and Marysville School District

A statistical analysis of skill development

CENTER · FOR EDUCATIONAL LEADERSHIP
UNIVERSITY OF WASHINGTON
College of Education

There was a time long ago when the most challenging role of the principal was to recognize and name all the students who came in the door. Principals are still expected to create a welcoming environment in their schools, but today their most challenging responsibility is to be instructional leaders. They need to be able to recognize and name the elements of instruction they observe in a classroom, for example, whether the teacher has clearly identified the purpose of the lesson and how student knowledge is being measured, and then based on what they observe, provide feedback to teachers and design professional development that helps teachers improve. No small feat.

Recognizing the importance of instructional leadership in improving classroom instruction, the Center for Educational Leadership (CEL) works in partnership with school districts to help school leaders increase their knowledge of what is being taught and their capacity to identify if it is being taught well. The scope of a CEL-district partnership is broader than just instructional leadership, touching all parts of the system (see Fig. 1, CEL Theory of Action) but as its name implies, the work is focused predominately on school leaders.

An examination of the theory of action of the Center for Educational Leadership (CEL) and research regarding its work in partnership with school districts is being conducted by the Center for the Study of Teaching and Policy at the University of Washington College of Education. This publication, the third in a series to summarize the research, presents findings from a statistical analysis by Michael Copland and Dina Blum and commentary from Stephen Fink, CEL Executive Director.

Researchers from the University of Washington describe this partnership work as an intervention—an attempt to increase the existing level of skill by providing professional development for school district administrators, principals, and district coaches. In a statistical analysis of a CEL intervention in two districts, they found evidence to suggest that school leaders did improve their ability to analyze instruction and provide feedback to teachers.

This publication provides a summary of:

- the nature of the partnership between CEL and two districts,
- the research design and methodology, and
- findings and recommendations for further study.

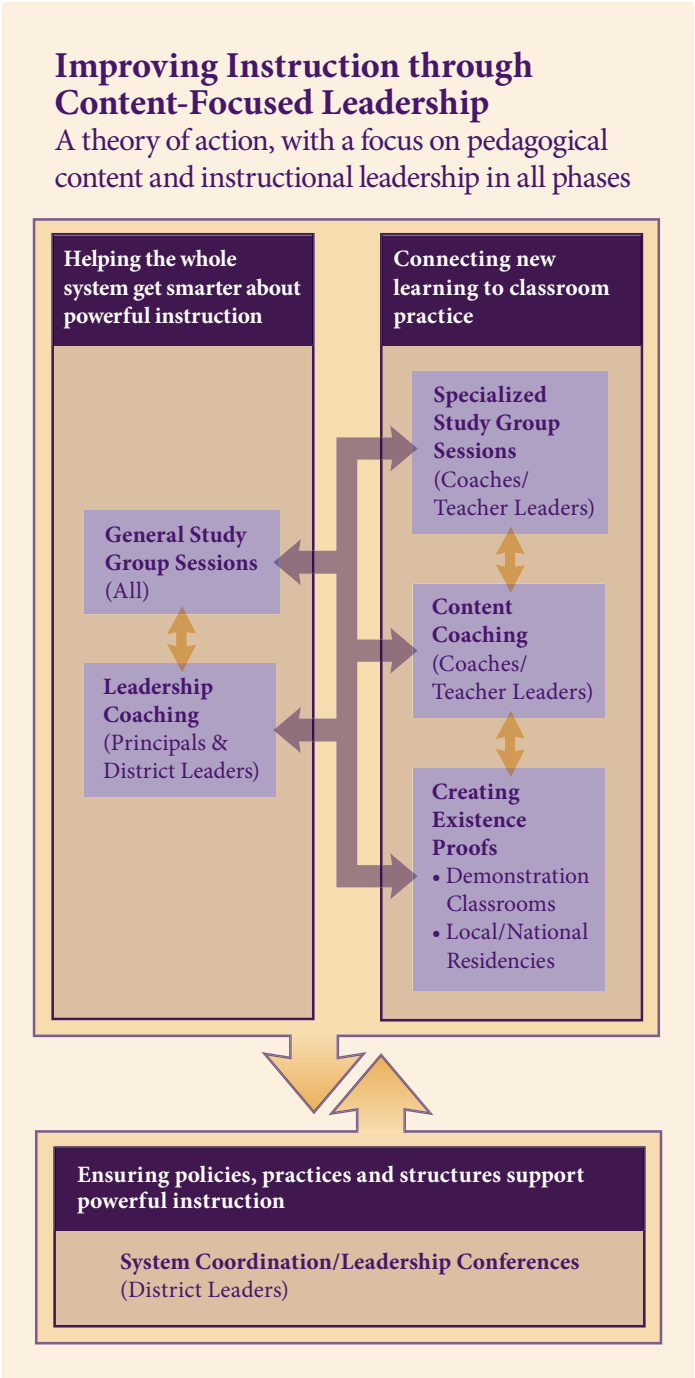
A partnership to improve instruction

The Center for Educational Leadership has formed a collaborative partnership with Norwalk-La Mirada (CA) School District and Marysville (WA) School District, providing expertise and consulting services with the specific intent of improving instructional leadership, and by extension, classroom instruction. One premise of this partnership is that the more an administrator or district coach knows about subject content and pedagogy (how to teach), the more expert his or her analysis of instruction is likely to be. With the development of this expertise, school leaders are better able to lead and guide professional development, align the resources they have to teacher development needs, engage in on-going problem solving, and build the instructional capacity of teachers.

Norwalk-La Mirada, a district whose 24,000 students are predominately Hispanic, has set a goal that 9 out of 10 of its students will read at grade level by the end of 2007. Marysville's student population of 12,000 is primarily white, but also includes a relatively high percentage of Native American students from the nearby Tulalip tribe. After experiencing the longest teacher strike in Washington state history, the board sought new district leadership in 2003 and endorsed a focus on improving literacy instruction.

As the researchers reported, "CEL's work with school and district leaders in these sites provided a ready-made laboratory for exploring instructional improvement work, and testing assumptions about whether and how leaders' skills with critical analysis and reflection on instructional practice improve over time."

Figure 1. CEL Theory of Action



A pre- and post-assessment to identify differences in skill levels

Given the commitment in these two districts to develop instructional leadership, the researchers posited a question and designed a study to learn whether the desired outcome was achieved through the CEL intervention: "Did the leaders' ability to critically analyze instruction and plan feedback for teachers deepen over time?"

To measure change—or the lack of it—in skill level, principals and district coaches in both districts analyzed a 15-minute video of classroom instruction in literacy at two points in time, Year 1 (2005) and Year 2 (2006). They then responded in writing to three questions and were scored on their answers:

1. What do you notice about teaching and learning in this classroom?
2. Given your response to Question #1, describe the follow-up conversation you would have with this teacher.
3. Imagine that the teacher you just observed is a member of your current school staff. What implications for professional development, if any, does this observation suggest?

In short, the research study quantified the answers with a 4-point scale, and used statistical models to determine if answers improved from one year to the next.

The scoring procedure itself provides insight into the complexity of teaching as well as observation of that teaching. Researchers created a rubric specifically for the study that reflected what CEL had been teaching through various components of the partnership, for example, how to know when teachers have engaged students in learning. They examined the responses for descriptions that referenced—or not—these categories of "instructional improvement" (see Table 1) and rated the responses on a 1-4 scale, from 1 for novice to 4 for expert (see Table 2). The rubric was created solely for the purpose of the research; the leaders who observed the lesson segment and provided responses to the three questions did not have prior knowledge of the rubric.

Each response was reviewed by two raters; when differences occurred in rating, a consensus score was determined after the two raters discussed their conclusions. Inter-rater reliability was .85 (with 1.0 as a perfect match).

Table 1. Categories and Sub-Dimensions of Instructional Improvement Rubric

Category	Sub-Dimensions
PURPOSE	(1) Standards
	(2) Teaching Point
STUDENT ENGAGEMENT	(3) Who's Doing the Work?
	(4) Student Engagement Strategies
	(5) Student Talk
CURRICULUM, PEDAGOGY, & ASSESSMENT	(6) Curriculum
	(7) Teaching Approach/Strategy
	(8) Scaffolds for Learning
	(9) Teaching Decisions
	(10) Assessment
CLASSROOM ENVIRONMENT & CULTURE	(11) Use of Physical Space
	(12) Classroom Routines
	(13) Classroom Work Culture

"The scoring procedure itself provides insight into the complexity of teaching as well as observation of that teaching."

Table 2. General Description: Expert/Novice Continuum

1 = Novice	2 = Emerging competence	3 = Developing expertise	4 = Expert
No mention of the phenomenon of interest, or Complete misconception about the phenomenon of interest	Noticing some of the structures of teaching (charts and room arrangements), but not the "whys" or "hows" underneath those structures with regard to the phenomenon of interest Non-analytical recounting of what transpired Superficial level of understanding of the phenomenon of interest Naming activities using the "right" language Does not discuss or elaborate on observation of activities	Appropriate mention of the phenomenon of interest Expresses wonder or questions teaching decisions or thinking behind teaching decisions Developing understanding that teaching decisions impact student outcomes, and how this occurs Noticing more subtle intentional teaching decisions; elements of structure and rationale for why Discusses/elaborates on notions raised in "emerging competence"	Demonstrates all of the markers of category three, plus additional subtleties of teaching and learning process related to the phenomenon of interest; identifies and acknowledges more layers of complexity Analytically unpacks teaching decisions Identifies complexities in connections between various elements of teaching and learning Considers teaching decisions in larger context (standards, unit of study)

In the report, the researchers provide a detailed description of each broad category, for example, purpose, and its sub-dimensions, in this case standards and teaching points. They illustrate responses related to each sub-dimension that would qualify as novice, emerging competence, developing expertise, and expert.

In the sub-dimension of teaching point, a novice response made no mention or recognition of teaching point, lesson purpose, or how the lesson connects to transferable skills, or showed that the observer had a misconception about teaching point. The response below illustrating a 3-response, developing expertise, discussed the lesson's teaching point, considered the connection to broader purpose (e.g., standards, outcomes, transferable knowledge or skill), and whether the lesson matched the stated purpose.

> *Teacher is providing feedback on how students are pointing out the symbolism….Students were able to identify author's use of silence to symbolize the chasm in communication…. Teacher asked students to consider the use of the cracker, Matzo, and what purpose it served in the story….The big idea in the lesson appeared to me to be: "Authors sometimes use objects, people, or events to symbolize bigger ideas and concepts."*

In the sub-dimension of curriculum, for example, if there were no mention of the curriculum used in the lesson, the leader received a novice score. The following response was scored as evidence of emerging competence:

> *I wanted to commend you on your text selection. You selected a text that lent itself to critical thinking and reflection about author's purpose and symbolism that seemed to be just right for the students I observed participating in the discussion.*

The raters would have given a score of expert if there had been "additional observations about how the text was a useful choice, further elaboration of what specifically made the text 'just right' for the students, or the formulation of questions designed to prompt the teacher's thinking about why the text was 'just right.'"

In the sub-dimension of student engagement, a novice response made no mention of the factors that encouraged or inhibited engagement. A response scored as emerging competence identified strategies to encourage participation such as group work, "turn and talk," or "think-pair-share." The following response was identified as developing expertise:

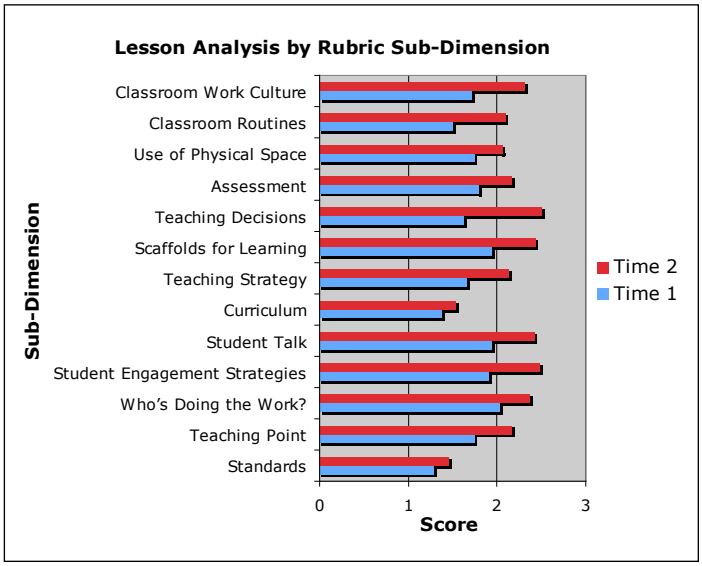
> *I would consider modeling for this teacher and others how to pace a discussion and embed time for pair/share in the discussion to allow for more students to participate while the thinking work is going on.*

"To the extent that principals can discern the differences in these sub-dimensions of instructional practice, they are more capable of orchestrating professional learning. The distinction between standards and teaching point is important, but in my experience, many principals and central office leaders conflate the two and simply expect teachers to have some perfunctory objective written on the board to satisfy the expectation of 'purpose.' A principal with a discerning eye may pick up this pattern in his/her building and then focus on doing some staff work or lesson study around teaching point." – Stephen Fink, CEL Executive Director

Both districts show statistically significant improvement

Using statistical tests, the researchers found significant differences in scores from Year 1 to Year 2 in Norwalk-La Mirada for all responses (grand mean), for responses compiled for each of the four broad categories, and for responses in 11 of the 13 sub-dimensions. The overall mean score for principals and coaches combined improved from 1.72 in Year 1 to 2.17 in Year 2.

Figure. 2 Norwalk-La Mirada Lesson Analysis by Rubric Sub-Dimension

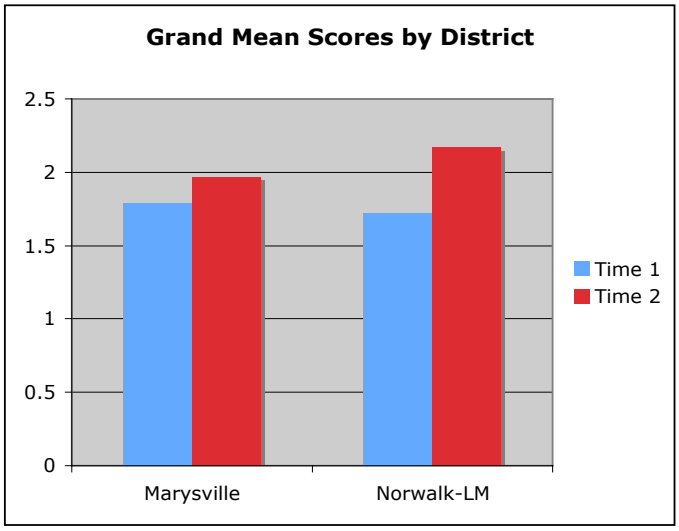


Comparing the scores of administrators with those of district coaches, researchers found that while both groups showed statistically significant improvement, the mean scores overall of the district coaches were slightly higher than the administrators and they made slightly larger gains over time.

"Principals and coaches in both districts improved in their ability to analyze instruction and plan comments to teachers."

The overall mean score also improved in Marysville, from 1.79 to 1.97. While there were improvements in each broad category, the changes did not reach significance at the level of alpha > .05. One of the 13 sub-dimensions, teaching point, showed statistically significant change from one year to the next.

Figure. 3 Grand Mean Scores: Norwalk-La Mirada and Marysville



Given a smaller sample of participants in Marysville, researchers did not compare scores of administrators and district coaches.

Conclusions and recommendations from the researchers

Concluding that principals and district coaches in both districts improved in their ability to analyze instruction and plan comments to teachers on their observations, the researchers nevertheless cautioned against making causal links between the intervention and the outcomes. "Because of the complex nature of instructional improvement work, and the myriad intervening variables that principals and district coaches are likely exposed to over the course of any given year, such causal claims are not warranted. However, the gains observed in the data are consistent with the instructional and leadership coaching interventions provided by the external partner organization, the Center for Educational Leadership…."

They offered two possibilities to explain the difference in results between Norwalk-La Mirada and Marysville:

- Norwalk-La Mirada had a much larger sample (n = 36) than Marysville (n= 8). Since the larger group size increases the variability of scores, differences in mean scores are more likely to show up there than in a smaller group.
- Norwalk-La Mirada had experienced two full years of CEL partnership work at the time of the final round of data collection, compared to one year in Marysville.

Given these two caveats and the findings from the data, researchers propose another round of data collection in Marysville, additional research to determine whether the gains evidenced in both districts continue, and efforts to learn how the leaders' new knowledge and skill play out in their work with teachers in the district.

9709 3rd Avenue NE
Suite 306
Seattle, WA 98115
Campus Mailbox: #358731

P: 206.221.6881
F: 206.221.6774
edlead@u.washington.edu
www.k-12leadership.org

CENTER · FOR
EDUCATIONAL
LEADERSHIP

UNIVERSITY OF WASHINGTON
College of Education