Getting Serious About School Reform:

Three Critical Commitments

Robert J. Marzano

Marzano & Associates, 2008

There is growing evidence that the actions of district and school leaders can have a substantial effect on student achievement (see Marzano, Waters, & McNulty, 2005; Marzano & Waters, in press). To illustrate, assume that a student at the 50th percentile in terms of her achievement in mathematics enrolls in a school whose principal is at the 98th percentile in terms of his leadership skills. Then assume that school is in a district whose superintendent is at the 98th percentile in terms of her leadership skills. One would predict that over time, the student's achievement would increase from the 50th percentile to the 67th percentile (Marzano & Waters, in press).

One might ask how such a relationship can exist since district and school leaders do not interact directly with students. The answer is that district and school leaders influence student achievement when they implement policies that directly affect what happens in classrooms. Stated differently, district and school leaders must ensure that specific interventions are enacted in every classroom in every school. Research and theory point to at least three critical inventions that should occur in every classroom, in every school throughout a district.

These three interventions might be thought of as critical "commitments" that must be made to students by district and school leaders. While making these commitments might appear to be a straightforward endeavor, they are not simple because they require that certain actions occur in

every classroom. Almost by definition this goes against the historical culture of districts and schools, which might be characterized as "loosely coupled"—individual schools within a district and individual teachers within a school operate in total autonomy and isolation. Over the decades, districts and schools have shown little interest in becoming "tightly coupled" organizations in which student achievement is the superordinate goal supported by uniform yet flexible behaviors in the classroom (Marzano & Waters, in press). The premise of this paper is that until districts and schools become tightly coupled regarding student achievement, they cannot be thought of as *serious* about school reform. Three critical commitments are described that as a group represent a serious commitment to reform.

Commitment #1: Develop a System of Individual Student Feedback at the District, School, and Classroom Levels

The first commitment addresses individual student feedback at the district, school, and classroom levels. The keyword in this intervention is *feedback.* Hattie's (1999; Hattie & Timperley, 2007) review of over 500 meta-analyses, involving 450,000 effect sizes from 180,000 studies representing approximately 20 to 30 million students, indicates that effective feedback is one of the most powerful influences on student achievement. When implemented, this intervention allows districts and schools to identify strengths and weaknesses of every student so that weaknesses may be addressed quickly and efficiently. The overall intent is that students do not move through the system from grade level to grade level with information and skill deficiencies that cumulatively impede their learning. This commitment typically plays out in three phases.

Phase 1: Track Student Progress on Selected Learning Goals Using a Formatively Based System of Assessment

The advantages of formative assessment have been abundantly clear ever since the publication of Black and Wiliam's (1998) meta-analysis of some 250 studies. Their overall conclusion was that when administered effectively formative assessments have the potential of enhancing student achievement by .7 standard deviations. This implies that a student at the 50th percentile in academic achievement might rise to the 76th percentile when exposed to effectively administered formative assessments. While teachers engage in formative assessments quite regularly, they

often do so in isolation, using their own idiosyncratic interpretations of effective formative assessment. To reap the full benefits of formative assessment, a districtwide approach must be established.

A districtwide approach begins by identifying specific instructional targets or "learning goals" for selected subject areas at each grade level. For example, after analyzing districtwide performance on a recently administered state test, a district might identify a specific learning goal for first-grade mathematics, first-grade reading, and first-grade writing. One mathematics learning goal for reading, writing, mathematics, and science might be identified for the entire year for each semester. Similar goals would be identified at each grade level. Individual schools within the district would be invited to identify additional grade level goals for their school.

Next, rubrics or scales would be developed for each goal at each grade level. Marzano (2006) has developed a generic scale that can be applied to all content areas (see Figure 1).

Figure 1: Generic Scale for District and Schoolwide Learning Goals

Score 4.0: In addition to score 3.0 performance, in-depth inferences and applications that go beyond what was taught.

Score 3.5: In addition to score 3.0 performance, partial success at inferences and applications that go beyond what was taught (score 4.0 elements).

Score 3.0: No major errors or omissions regarding any of the information and/or processes (simple or complex) that were explicitly taught.

Score 2.5: No major errors or omissions regarding the simpler details and processes (score 2.0 elements) and partial knowledge of the more complex ideas and processes (score 3.0 elements).

Score 2.0: No major errors or omissions regarding the simpler details and processes (score 2.0 elements) but major errors or omissions regarding the more complex ideas and processes (score 3.0 elements).

Score 1.5: Partial knowledge of the simpler details and processes (score 2.0 elements) but major errors or omissions regarding the more complex ideas and

processes (score 3.0 elements).

Score 1.0: With help, a partial understanding of some of the simpler details and processes (Score 2.0 elements) and some of the more complex ideas and processes (Score 3.0 elements).

Score 0.5: With help, a partial understanding of some of the simpler details and processes (score 2.0 elements) but not the more complex ideas and processes (score 3.0 elements).

Score 0.0: Even with help, no understanding or skill demonstrated.

To illustrate how this scale is used, assume that a district identified the following instructional goal for eighth-grade science.

Students will understand:

- Distinctions between asexual and sexual reproduction regarding risk of mutation, energy requirements, similarities of offspring to parents, and the processes involved
- The impact of heredity on organisms regarding specific traits of organisms, diseases, and genetic disorders

That goal would be translated into a scale using the model in Figure 1. This is depicted in Figure 2.

Figure 2: Common Scale for Eighth-Grade Science Goal

Score 4.0: In addition to score 3.0, in-depth inferences and applications that go beyond what was taught, such as:

- **Describing how a genetic disorder (for example, cystic fibrosis) can be passed from parents to offspring when the parents are healthy**

Score 3.5: In addition to score 3.0 performance, in-depth inferences and applications with partial success.

Score 3.0: While engaged in tasks that address principles of heredity, the student demonstrates an understanding of important information such as:

- **Distinctions between asexual and sexual reproduction (risk of mutation, energy requirements, similarity of offspring to parent, processes involved)** ***(for example, explaining how asexual and sexual reproduction differ in their impact on potential mutation of offspring, such as by describing which type of reproduction has a greater risk of mutation and why the risk is greater)***
- **The impact of heredity on organisms (traits, diseases, genetic disorders)** ***(for example, describing how a trait such as body type can affect the lives of the members of a family across generations)***

The student makes no major errors or omissions.

Score 2.5: No major errors or omissions regarding the score 2.0 elements and partial knowledge of the score 3.0 elements.

Score 2.0: No major errors or omissions regarding the simpler details and processes such as:

- **Recognizing and recalling specific terminology, such as** ***egg, sperm, genetic mutation, offspring, organism, reproduction, heritable characteristics***
- **Recognizing and recalling isolated details, such as:**
  - **Half the genes come from each parent in sexual reproduction.**
  - **Heritable characteristics determine an organism's likelihood to survive and reproduce.**

However, the student exhibits major errors or omissions with score 3.0 elements.

Score 1.5: Partial knowledge of the score 2.0 elements, but major errors or omissions regarding the score 3.0 elements.

Score 1.0: With help, a partial understanding of some of the score 2.0 elements and some of the score 3.0 elements.

Score 0.5: With help, a partial understanding of some of the score 2.0 elements,

but not the score 3.0 elements.

Score 0.0: Even with help, no understanding or skill demonstrated.

With this scale that is specific to eighth-grade science, teachers could readily develop formal and informal formative assessments. Although teachers would be designing their own assessments, each assessment would be scored with the same scale, allowing for comparability of student progress from teacher to teacher. A student with a score of 2.5 on an assessment for a specific learning goal designed by one teacher would be comparable to a score of 2.5 on an assessment designed by another teacher for that same learning goal. This system does not preclude the use of common assessments. Common assessments would be designed and scored using the common scale. Thus the district would have a comprehensive yet flexible system of formative assessments to track students' progress in a way that is comparable from teacher to teacher and school to school.

Finally, using DuFour, Eaker, and DuFour's (2005) conception of professional learning communities, teachers would meet periodically to examine the achievement of their students on learning goals assessed using the common scale. Discussions at the meetings would focus on identifying instructional strategies that produce the greatest gains in student learning.

Phase II: Design Learning Goals in All Subject Areas and Redesign Report Cards

Phase I allows a district to keep track of student progress on selected learning goals. Ultimately, a district must be able to keep track of a comprehensive set of learning goals for each subject area as opposed to a few selected learning goals for a few subjects. This means that a district must reconstitute state standards documents (see Marzano & Haystead, 2008). One aspect of reconstituting standards documents involves trimming the amount of content students are expected to learn and teachers are expected to teach. This is necessary because research has shown that if all the content in current standards documents were taught, schools would have to add about 70% more time to the school year (Marzano, Kendall, & Gaddy, 1999).

Along with trimming the content, learning goals at each grade level would be stated in scale format as depicted earlier in Figure 2. Marzano and Haystead (2008) have determined that no more than 15 learning goals should be identified for any given grade level. These learning goals are typically organized into broader categories called *strands.* Some districts and schools refer to refer to learning goals as *reporting topics***.** Figure 3 depicts sample strands and reporting topics as described by Marzano (2007b).

Figure 3: Sample Strands and Reporting Topics

Language Arts

Reading:

1. Word recognition and vocabulary
2. Reading comprehension
3. Literary Analysis

Writing:

4. Spelling
5. Language mechanics and conventions
6. Research and technology
7. Evaluation and revision

Listening and Speaking:

8. Listening comprehension
9. Analysis and evaluation of oral media
10. Speaking applications

Mathematics

Numbers and Operations:

1. Number sense and number systems
2. Operations and estimation

Computation:

3. Addition and subtraction
4. Multiplication and division

Algebra and Functions:

5. Patterns, relations, and functions
6. Algebraic representations and mathematical models

Geometry:

7. Lines, angles, and geometric objects
8. Transformations, congruency, and similarity

Measurement:

9. Measurement systems
10. Perimeter, area, and volume

Data Analysis and Probability:

11. Data organization and interpretation
12. Probability

Science

Nature of Science:

1. Nature of scientific knowledge and inquiry
2. Scientific enterprise

Physical Sciences:

3. Structure and properties of matter
4. Sources and properties of energy
5. Forces and motion

Life Sciences:

6. Biological evolution and diversity of life
7. Principles of heredity and related concepts
8. Structure and function of cells and organisms
9. Relationships among organisms and their physical environment

Earth and Space Sciences:

10. Atmospheric processes and the water cycle
11. Composition and structure of the Earth
12. Composition and structure of the Universe and the Earth's place in it

Social Studies

Citizenship, Government, and Democracy:

1. Rights, responsibilities, and participation in the political process
2. The U.S. and state constitutions
3. The civil and criminal legal systems

Culture and Cultural Diversity:

4. The nature and influence of culture

Economics:

5. The nature and function of economic systems
6. Economics throughout the world
7. Personal economics

History:

8. Significant individuals and events
9. Current events and the modern world

Geography:

10. Spatial thinking and the use of charts, maps, and graphs

With strands and learning goals in place, report cards can be redesigned, as depicted in Figure 4.

Figure 4: Sample Report Card

Name: John Mark

Address: 123 Some Street

City: Anytown, CO 80000

Grade Level: 4

Homeroom: Ms. Smith

Language Arts	2.46	C	Participation	3.40	A
Mathematics	2.50	B	Work Completion	2.90	B
Science	2.20	C	Behavior	3.40	A
Social Studies	3.10	A	Working in Groups	2.70	B
Art	3.00	A			

Language Arts

Reading:

Word Recognition and Vocabulary 2.5

Reading for Main Idea 1.5

Literary Analysis 2.0

Writing:

Language Conventions 3.5

Organization and Focus	2.5	
Research and Technology	1.0	
Evaluation and Revision	2.5	
Writing Applications	3.0	
Listening and Speaking:		
Comprehension	3.0	
Organization and Delivery	3.0	
Analysis and Evaluation of Oral Media	2.5	
Speaking Applications	2.5	
Life Skills:		
Participation	4.0	
Work Completion	3.5	

Behavior	3.5	
Working in Groups	3.0	
Average for Language Arts	2.46	
Mathematics		
Number Systems	3.5	
Estimation	3.0	
Addition/Subtraction	2.5	
Multiplication/Division	2.5	
Ratio/Proportion/Percent	1.0	
Life Skills:		
Participation	4.0	
Work Completion	2.0	

Behavior	3.5	
Working in Groups	2.0	
Average for Mathematics	2.50	

Science		
Matter and Energy	2.0	
Forces of Nature	2.5	
Diversity of Life	1.5	
Human Identity	3.5	
Interdependence of Life	1.5	
Life Skills:		
Participation	3.0	

Work Completion	1.5	
Behavior	2.5	
Working in Groups	1.0	
Average for Science	2.20	

Social Studies		
The Influence of Culture	3.5	
Current Events	3.0	
Personal Responsibility	4.0	
Government Representation	3.5	
Human and Civil Rights	1.5	
Life Skills:		
Participation	3.5	

Work Completion	3.5	
Behavior	3.5	
Working in Groups	4.0	
Average for Social Studies	3.10	

Art

Purposes of Art	3.5	
Art Skills	3.0	
Art and Culture	2.5	

Life Skills:

Participation	2.5	
Work Completion	4.0	

Behavior	4.0	
Working in Groups	3.5	
Average for Art	3.00	

The report card in Figure 4 is traditional in the sense that it provides overall grades. This is done by combining the final status for the reporting topics in each subject area into a weighted or unweighted average and then converting that average to a traditional letter grade using a conversion like the following: 3.00 to 4.00 = A; 2.50 to 2.99 = B; 2.00 to 2.49 = C; 1.50 to 1.99 = D; below 1.50 = F. In addition to this overall grade, student status on each reporting topic is depicted. To illustrate, consider the topic of reading for main idea within the general category of reading. The bar graph for reading in Figure 4 shows that the student's final score was 3.0, indicating the student's status at the end of the grading period.

Of course, this is only one example of the various types of report cards that can be designed. A district might elect to have no overall traditional grade and simply report status on individual reporting topics. Additionally, a district might elect to have report cards depict knowledge gain for each reporting topic or learning goal in addition to reporting final status. (For a detailed discussion of the various types of report cards, see Marzano, 2006).

Phase III: Implement the New Report Cards in a Staged Fashion

During Phase II, report cards are designed but not implemented. During Phase III, the district implements the new report cards. While a case could be made that new report cards should be introduced systemwide at one time, an equally logical perspective is to implement in a staged fashion. For example, a district might implement the new report cards at the elementary level first. The next year, the district might implement the new report cards at the middle school level.

The third year, the district would implement the new report cards at the high school level. Assuming that Phases I and II take 1 year each, the entire process of designing and implementing this first critical commitment would take 5 years.

Commitment #2: Ensure Effective Teaching in Every Classroom

The second commitment addresses effective teaching. Although it is probably true that a district cannot *ensure* effective teaching in every classroom, it can implement a system that ensures a districtwide emphasis on monitoring and enhancing the effectiveness of teaching in every classroom. This commitment involves four phases.

Phase 1: Systematically Explore and Examine Effective Pedagogy and Develop a Model or "Language" of Instruction

Over the years, a number of models of effective pedagogy have been proposed (for example, Hunter, 1984). While a case can be made that a district or school should simply adopt a model, a case can also be made that "off-the-shelf" interventions are typically short lived in K–12 education (Cuban, 1987). One alternative to adopting an instructional model wholesale is to use action research to develop a local or district or school approach.

The concept of action research has become quite popular in the last few decades. Nolen and Putten (2007) note that action research was first introduced as a methodology in education research in the mid-1950s. They explain that it "surfaced in response to the growing need for more relevant and practical knowledge in the social sciences: It bridged the gap between academic research and day-to-day applications" (p. 401).

For the purposes discussed here, action research begins with the identification of specific instructional techniques that are to be studied. This typically means selecting strategies from existing lists of effective practices. For example, relative to instructional strategies, Marzano, Pickering, and Pollock (2001) identified the following nine instructional strategies:

1. Identifying similarities and differences

2. Summarizing and note-taking

3. Reinforcing effort and providing recognition

4. Homework and practice

5. Nonlinguistic representations

6. Cooperative learning

7. Setting objectives and providing feedback

8. Generating and testing hypotheses

9. Cues, questions, and advance organizers

Relative to classroom management strategies, Marzano, Pickering, and Marzano (2003) identified the following four areas:

1. Rules and procedures

2. Disciplinary interventions

3. Teacher-student relationships

4. Teacher mental set

Other similar lists of effective strategies have been developed by Good and Brophy (2003) and Mayer (2003).

Once a reference list of strategies has been identified, teachers throughout a school or district can conduct action research projects on a voluntary basis. Action research projects can be quite informal or formal. At an informal level, teachers might simply try strategies in their classrooms and record their impressions of how well they worked. At a more formal and more rigorous level, teachers can design and carry out studies involving experimental classes (classes in which a specific strategy is employed) and control classes (classes in which the selected strategy is not employed). To date, Marzano & Associates has been involved in over 100 action research projects that employed experimental/control classes and controlled for previous knowledge using pretests as covariates (see, for example, Marzano & Associates, 2005). Results from these studies are reported in Figure 5.

Figure 5: Distribution of 113 Effect Sizes (Standardized Mean Difference)

Mean	.39
Median	.28
Range	6.67
10th percentile	-.43
20th percentile	-.17
25th percentile	-.06
30th percentile	.06
40th percentile	.15
50th percentile	.28

60th percentile	.33
70th percentile	.61
75th percentile	.70
80th percentile	.90
90th percentile	1.44

In Figure 5, effect sizes are reported in terms of standardized mean differences. Thus an effect size of .25 (let's say) means that the average score in a class in which a specific strategy was used would be expected to be about 10 percentile points larger than average score in a class where the strategy was not used. The average effect size in Figure 5 is .39, which implies a 15 percentile point differential between the average score in a class where a specific strategy was used and the average score in a class where it was not. Of particular interest in these studies is the fact that the vast majority of the 113 teachers either participated in a one-day or half-day in-service professional development workshop regarding specific instructional strategies, read a brief description of a specific instructional strategy, or both. An average effect size of .39 or 15 percentile points can be considerable noteworthy under these conditions.

Once valued instructional strategies have been studied via action research, a school or district is in a position to design a model or "language" of instruction. An instructional model should not be misconstrued as an attempt to constrain teachers to one particular approach to teaching. Rather, it should be interpreted as a necessary vehicle for communication between teachers. In effect, the model should constitute an agreed-upon way of describing and discussing effective teaching. This idea has been espoused by many. For example, the importance of a common language is addressed implicitly and explicitly by those who promote the importance of professional learning communities (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006).

In *The Art & Science of Teaching*, Marzano (2007a) has offered the approach depicted in Figure 6.

Figure 6: Instructional Design Questions

1. What will I do to establish and communicate learning goals, track student progress, and celebrate success?

2. What will I do to help students effectively interact with new knowledge?

3. What will I do to help students practice and deepen their understanding of new knowledge?

4. What will I do to help students generate and test hypotheses about new knowledge?

5. What will I do to engage students?

6. What will I do to establish or maintain classroom rules and procedures?

7. What will I do to recognize and acknowledge adherence and lack of adherence to classroom rules and procedures?

8. What will I do to establish and maintain effective relationships with students?

9. What will I do to communicate high expectations for all students?

10. What will I do to develop effective lessons organized into a cohesive unit?

Design questions such as those in Figure 4 have an advantage over models that focus on lessons (such as Hunter, 1984) in that they allow a great deal of flexibility for teachers in terms of the day-to-day practice of teaching but are specific enough that they allow for detailed discussion of the teaching/learning process.

Phase II: Have Teachers Systematically Interact Using the Model or Language of Instruction

A model of instruction is powerful only if used as a vehicle for communication—as the basis for conversations about effective teaching. In a school with a culture of effective instruction many of these conversations will occur quite naturally. While naturally occurring interactions should be supported, it is also important to provide a structure for these interactions. Dimmock (2000) notes that providing teachers with the time and space to interact about instruction is critical to effective interaction; however, time and space are not sufficient. A format and structure for such interactions should be developed.

To illustrate, a district might use "late starts" on a monthly basis. During late start meetings, teachers meet in small grade-level or subject-matter teams to discuss instructional issues. Between late start meetings, teachers are asked to record their reactions to instructional techniques they have tried from the district model. This recordkeeping is kept to a minimum; teachers simply spend a few minutes after a particular lesson in which they tried a strategy recording their perceptions of the effectiveness of the strategy. During their late start meetings, teachers discuss their recorded observations using the following protocol:

- Describe the strategy or strategies you tried.
- Describe its effect on student learning and the evidence for your conclusions.

- Describe what you did.
- Describe areas for improvement on your part.

Teacher monthly meetings might be coupled with an examination of students' progress on specific learning goals gleaned from the first commitment above. That is, if multiple teachers have been focusing on a specific learning goal within a grade level or for a given course, they correlate their discussions of the effectiveness of specific instructional strategies with student progress on the common scale for a specific learning goal.

Phase III: Have Teachers Observe Master Teachers Applying Instructional Strategies

Teachers systematically talking about instruction will go a long way to creating a culture of effective teaching. However, nothing will put effective pedagogy in the spotlight as well as teachers observing teachers. Louis, Kruse, & Associates (1995) note that ultimately professional learning communities (PLCs) must foster the "deprivatization of practice." This is perhaps one the most difficult aspects of PLCs to implement. In his book *A Place Called School,* which summarized data from 1,350 elementary and secondary teachers, Goodlad (1984) noted that teachers generally report that they would like to observe others: "Approximately three quarters of our sample at all levels of schooling indicted that they would like to observe other teachers at work" (p. 188).

This phase necessarily begins with the identification of "master teachers." It is important to note that a master teacher is defined as one who produces substantial gains in student learning. This is in contrast to defining a master teacher as one who employs specific instructional strategies. Although this might seem counterintuitive, it has a strong logic. Given the complexity of the teaching/learning process, it is safe to say that no model of instruction or set of instructional strategies could completely define effective teaching. This sentiment has been expressed directly or indirectly by many researchers and theorists (Willms, 1992; Reynolds & Teddlie, 2000; Berliner, 1986). Different teachers employing the same instructional techniques might produce

very different results in student learning. Consequently, overall effectiveness in teaching must be defined in terms of the one indisputable criterion—student learning.

Once master teachers have been identified using the criterion of student learning, each master teacher's strengths can be identified in terms of the model of instruction that has been designed by the school or district. To illustrate, assume that a district is using the 10 design questions depicted in Figure 4. Within a particular school, a master teacher might be identified who demonstrates skill at questions 1 and 3. Another master teacher in another school might demonstrate expertise at questions 2 and 5 and so on. By definition, both master teachers produce consistent learning in their classrooms, but each has different strengths.

On a voluntary basis, teachers would then sign up to observe master teachers for specific design questions. For example, if a teacher wanted to observe an expert in question 5—student engagement—she would seek out one of the district experts on this issue. Ideally, the expert teacher for design question 5 would also visit the classroom of the teacher seeking assistance.

Phase IV: Monitor the Effectiveness of Individual Teaching Styles

The final phase of the second critical commitment involves monitoring instruction districtwide or schoolwide. This means that every teacher is provided with feedback regarding the effectiveness of their instruction with the intent of capitalizing on strengths and improving on weaknesses. One important reminder is useful here: The criterion for effective teaching should be s*tudent learning* as opposed to the rigid use of strategies identified in the instructional model. The instructional model is a means to an end, not an end in itself. As Fullan (2001) explains, the purpose of teacher observation is to produce, through interaction, shared knowledge that teachers can apply to address real-world issues in their classrooms. Hord (1997) echoes these comments, noting that shared knowledge regarding instruction should translate into practical tools that can be used by teachers to enhance student achievement.

While it is true that teachers should be allowed flexibility in the instructional strategies they employ, it is also true that all teachers should be expected to produce "learning" in their

classrooms. Stated differently, teachers should be allowed to exhibit wide variation (that is, have different profiles) as to various aspects of the instructional model they emphasize. However, there should be no variation in *expectations* about student learning from teacher to teacher.

To this end, it is recommended that data are systematically collected on students as well as teachers. For teachers, data would be collected on the extent to which they employ elements of the instructional model. Such data might be collected via supervisor observations and teacher self-report. A rubric like that depicted in Figure 1 should be developed for each element of the instructional model.

Figure 7: Teacher Rubric for Design Question 1—What will I do to establish and communicate learning goals, track student progress, and celebrate success?

Score 4.0: In addition to score 3.0 behaviors, adaptations that enhance students' learning.

Score 3.5: In addition to score 3.0 behaviors, partial success with adaptations that enhance students' learning.

Score 3.0: While engaged in classroom activities that involve establishing and communicating learning goals, tracking student progress, and celebrating success, the teacher makes no major errors or omissions regarding the following behaviors:

- **Presents students with a clearly defined scale or rubric for each learning goal**
- **Allows students to identify their own learning goals in addition to those presented to them**
- **Designs and administers formative assessments for each learning goal**
- **Displays progress on learning goals for the whole class and facilitates students tracking their own progress**
- **Recognizes individual student status and progress as well as that of the whole class**

Score 2.5: No major errors or omissions regarding the simpler behaviors (score 2.0 performance) and partial success at the more complex behaviors (score 3.0 performance).

Score 2.0: No major errors or omissions regarding the following simpler behaviors:

- **Makes a distinction between learning goals and learning activities**
- **Presents learning goals, but does not design a scale for each**
- **Designs and administers assessments for each learning goal, but does not use a formative system**
- **Tracks student progress, but does not facilitate students tracking their own progress or does not display progress for the whole class**
- **Recognizes and celebrates individual status and progress or group status and progress, but not both**

However, the teacher exhibits major errors or omissions regarding the more complex behaviors (score 3.0 performance).

Score 1.5: Partial success at the simpler behaviors (score 2.0 performance), but major errors or omissions regarding the more complex behaviors (score 3.0 performance).

Score 1.0: With help, partial success at some of the simpler behaviors (score 2.0 performance) and some of the more complex behaviors (score 3.0 performance).

Score 0.5: With help, partial success some of the simpler behaviors (score 2.0 performance), but not the more complex behaviors (score 3.0 performance).

Score 0.0: Even with help, no success with the score 2.0 or 3.0 behaviors.

Using rubrics such as that in Figure 7, a profile of each teacher can be compiled through teacher self-report and observations by supervisors. Regarding self-reports, teachers can rate themselves on a systematic basis and compile these ratings. Additionally, supervisors can make systematic observations of teachers. These two sources of data can be combined to construct a profile for

each teacher regarding their use of the instructional model. This is shown in the first 11 columns of Figure 8.

Figure 8: Summary Data for Teacher Profiles

Teacher	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Pre/post	Engagement	Student Learning
Teacher 1	2.5	2.5	3.0	2.0	2.0	3.0	4.0	3.5	3.0	2.5	.7	2.1	3.3
Supervisor			2.5			3.0			2.5				
Teacher 2	2.0	3.0	2.5	3.0	3.5	2.5	3.0	2.0	2.5	3.0	1.3	3.2	3.0
Supervisor		2.0			2.5	2.5							
School Average	2.4	2.4	3.1	2.7	3.1	2.6	2.8	2.9	3.0	3.2	.5	2.3	2.5
District Average	2.6	2.8	2.8	3.2	2.7	2.4	2.8	2.7	3.0	3.2	.4	2.7	2.4

A report like that in Figure 8 would be generated for each school within a district. Each pair of rows in Figure 8 represents the self-report and supervisor report data for a specific teacher regarding the 10 design questions for the model shown in Figure 4. To illustrate, consider the first two rows in Figure 8. The teacher has provided self-report scores for each of the 10 design questions. The teacher simply rated himself on each question using rubrics like that shown in Figure 7. Supervisor ratings are right below teacher ratings. Note that supervisor ratings are not

reported for all 10 design questions. This is because a supervisor would not have the time to make valid observations on all 10 design questions in a single year. Consequently, in a given year and in consultation with a given teacher, a supervisor would identify a few design questions to observe. Over the years, scores on all design questions would be obtained from supervisors. Also note that teacher names are not used. A composite report like that depicted in Figure 8 lists all teachers, but only individual teachers and supervisors know the identity of specific teachers. Finally, note that the last two rows for each design question report district and school averages respectively. This allows comparison of individual teacher profile data with that for the entire school and the entire district.

To complete the profile for each teacher, student data must be collected. Three types of student data are recommended. The first is pre-test/post-test also data from a specific unit of instruction. These pre-test/post-test scores should all use the same metric. The 0 through 4 scale in Figure 1 is recommended so that assessments from teacher to teacher follow the same metric. The pre-tests and post-tests can either be teacher-made assessments or common assessments designed by the district. Another type of data are student self-report engagement data. Again, a 0 through 4 scale is recommended. The final type of data is student self-report on their learning, using a 0 through 4 scale. The last two columns in Figure 8 show class averages for each teacher on the three types of student self-report data. As before, the last two rows contain school and district averages for the student data.

It is certainly not the case that every teacher should be expected to meet or exceed district or school averages in all measures. Each teacher is unique in his or her instructional profile. However, comparison between individual teacher profiles and school or district averages should form the basis for discussion between teachers and supervisors. More specifically, individual teachers in consultation with their supervisors should identify specific goals for improvement. The focal point of such deliberation should always be student learning and engagement. An individual teacher might set a goal of raising her pre-test/post-test achievement gain from .5 to .6 and raising the average level of student engagement by one half a scale point by the end of the year. The teacher might elect to focus on one specific instructional design question to achieve

these goals. For example, after examining her instructional profile as compared to the profiles of others in the district, the teacher might choose to focus on design question 1 of the model throughout the year. That selection might be made because the teacher notes that her scores on that design question are significantly below the school or district average.

Assuming that each phase takes 1 year, this second critical commitment can be designed and implemented in 4 years.

Commitment #3: Build Background Knowledge for All Students (Particularly Those With Educationally Challenging Backgrounds)

The third commitment addresses the academic background knowledge of students. Numerous studies have confirmed the relationship between background knowledge and achievement (Nagy, Anderson, & Herman, 1987; Dochy, Segers, & Buehl, 1999; Tobias, 1994; Schiefele & Krapp, 1996; Tamir, 1996; Boulanger, 1981). The average correlation reported in these studies between a person's background knowledge for a given topic and the extent to which that person learns new information regarding that topic is .66. This is a remarkably large correlation within the field of education.

This paints a compelling picture as to the importance of *academic* background knowledge to the academic success of students. It is important to acknowledge use of the qualifier *academic.* Two students might have an equal amount of background knowledge. However, one student's background knowledge might relate to traditional school subjects such as mathematics, science, history, and the like. The other student's equally large store of background knowledge might be about nonacademic topics such as the best set of trains to take in the subway to get downtown during rush hour, the place to stand in the subway car that provides the most ventilation on a hot summer day, the best food to bring on the subway in terms of ease of consumption, and so on. The importance of one type of background knowledge over another is strictly a function of context (Becker, 1977; Greenfield, 1998).

Marzano (2004) has demonstrated that vocabulary knowledge and background knowledge are for all practical purposes synonymous. Nagy and Herman (1984) found a consistent difference in vocabulary knowledge between students at different family income levels. They estimated a 4,700 word difference in vocabulary knowledge between high and low SES (socioeconomic status) students. Similarly, they estimated that mid-SES first graders know about 50% more words than do low-SES first graders. Graves and Slater (1987) found that first graders from higher-income backgrounds had about double the vocabulary size of those from lower-income backgrounds. Hart and Risley (1995) found that the differences in vocabulary development due to family status start at a very early age. They computed the correlation between vocabulary knowledge and family income to be .65—again, a very large correlation in the social sciences.

The relationship between vocabulary knowledge and academic achievement is also well established. As early as 1941, researchers estimated that for students in grades 4 through 12, there was about a 6,000 word gap between students at the 25th and 50th percentiles on standardized tests (see Nagy & Herman, 1984). Using a more advanced method of calculating vocabulary size, Nagy and Herman (1984) estimated the difference to be anywhere between 4,500 and 5,400 words for low- versus high-achieving students.

Given these findings from the research literature, the third commitment districts should make is to enhance students' academic background knowledge through direct instruction in specific academic terms. There are two phases involved in a districtwide approach to increasing academic background. These have been described in depth in Marzano (2004) and Marzano and Pickering (2006).

Phase I: Identify Academic Terms in Language Arts, Mathematics, Science, and Social Studies to Be Taught at Each Grade Level

During this first phase, a district identifies academic terms in language arts, mathematics, science, and social studies. (Academic terms for other subject areas may also be identified.) To illustrate, the following typifies the types of mathematics terms that might be identified at third grade:

Angle

Area

Average

Bar graph

Congruent

Difference

Estimation

Hexagon

Length

Width

These terms are conceptual in nature. It is recommended that about 30 terms are identified for each grade level for each subject area. In this way teachers will have time to teach other terms of their own choosing. Specifically, it has been estimated that teachers try to teach as many as 400 terms per year (see Marzano, 2004). If 30 terms only are identified per grade level for each of four subject areas, then a teacher in a self contained classroom would be required to teach 120 district-identified terms, leaving time for the teacher to address 280 terms of her own choosing.

Phase II: Implement the Academic Vocabulary Program Districtwide Using a Common Approach to Instruction

Once academic terms have been identified, a program of direct instruction in the academic terms can be implemented districtwide. As much as possible, a common approach to instruction should be used. Marzano (2004: Marzano & Pickering, 2005) has recommended the six-step process in Figure 9.

Figure 9: A Six-Step Process for Direct Vocabulary Instruction

Step 1: Provide a description, explanation, or example of the new term.

Step 2: Ask students to restate the description, explanation, or example in their own words.

Step 3: Ask students to construct a picture, symbol, or graphic representing the term.

Step 4: Engage students periodically in activities that help them add to their knowledge of the terms in their notebooks.

Step 5: Periodically ask students to discuss the terms with one another.

Step 6: Involve students periodically in games that allow them to play with terms.

This process allows for instructional flexibility from teacher to teacher but also provides students with an approach to learning new terms that is common regardless of the teacher or the term.

Assuming that Phases I and II will take one half year each, this third critical commitment can be designed and implemented in 1 year.

Conclusions

The actions of districts and schools can have a profound impact on student achievement. Yet historically, districts and schools have been so loosely coupled that they have had little influence on what occurs in individual classrooms and consequently have had little influence on student achievement. This paper has outlined three critical commitments that districts and schools can make. Implementing these critical commitments constitutes a concerted effort to be serious about school reform.

References

Becker, W. C. (1977). Teaching reading and language to the disadvantaged—What we have learned from field research. *Harvard Educational Review, 47*(4), 518–543.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*(7), 5–13.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–75.

Boulanger, D. F. (1981). Ability and science learning. *Journal of Research in Science Teaching,18*(2), 113–121.

Cuban, L. (1987, July). *Constancy and change in schools (1880s to the present).* Paper presented at the Conference on Restructuring Education, Keystone, CO.

Dimmock, C. (2000). *Designing the learning-centred school: A cross cultural perspective.* London: Falmer Press.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relationship between assessment practicesand outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145–186.

DuFour, R., Eaker, R., & DuFour, R. (Eds.). (2005). *On common ground: The power of professional learning communities.* Bloomington, IN: Solution Tree (formerly National Educational Service).

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York andLondon: Teachers College Press and RoutledgeFalmer.

Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms* (9th ed.). Boston: Allyn & Bacon.

Goodlad, J. I. (1984). *A place called school.* New York: McGraw-Hill.

Graves, M. F., & Slater, W. H. (1987, April). *The development of reading vocabularies on ruraldisadvantaged students, inner-city disadvantaged students, and middle-class suburban students.* Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures.* (pp. 81–123). Washington, DC: American Psychological Association.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

Hattie, J. A. (1999, June). *Influences on student learning* (Inaugural professional address. University of Auckland, New Zealand). Accessed at www.arts.auckland.ac.nz/staff/index.cfm?P=8650.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research.77*(1), 81–112.

Hord, S.M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement.* Austin, TX: Southwest Educational Development Laboratory.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. Hosford (Ed.), *Using what we know about teaching* (pp. 169–192). Alexandria, VA: Association for Supervision and Curriculum Development.

Louis, K.S., Kruse, S.D., & Associates. (1995). *Professionalism and community: Perspectives on reforming urban schools.* Thousand Oaks, CA: Corwin.

Marzano & Associates. (2005). *Blue Springs School District instructional strategies project: Final report.* Centennial, CO: Author.

Marzano, R. J. (2004). *Building background knowledge for academic achievement: What works in schools.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2006). *Classroom assessment and grading that work.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007a). *The art and science of teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007b). Designing a comprehensive approach to classroom assessment. In D. Reeves (Ed.). *Ahead of the curve: The power of assessments to transform teaching and learning* (pp. 103–126). Bloomington, IN: Solution Tree.

Marzano, R. J., & Haystead, M. W. (2008). *Making standards useful in the classroom.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Kendall, J. S., & Gaddy, B. B. (1999). *Essential knowledge: The debate over what American students should know.* Aurora, CO: Mid-continent Research for Education and Learning.

Marzano, R. J., & Pickering, D. J. (2005).*Building academic vocabulary: Teacher's manual.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Marzano, J. S. (2003). *Classroom management that works: Research-based strategies for every teacher.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Waters, T. (in press). *From the boardroom to the classroom: District leadership that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: Fromresearch to practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mayer, R.E. (2003). *Learning and instruction.* Upper Saddle River, NJ: Merrill, Prentice Hall.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meaning from context during normal reading. *American Educational Research Journal. 24*(2), 237–270.

Nagy, W. E., & Herman, P. A. (1984). *Limitations of vocabulary instruction* (Tech. Rep. No. 326). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Document Reproduction Service No. ED248498)

Nolen, A. L., & Putten, J. V. (2007). Action research in education: Addressing gaps in ethical principles and practices. *Educational Researcher, 36*(7), 401–407.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*(1), 160–164.

Reynolds, D., & Teddlie, C. (with Hopkins, D., & Stringfield, S.). (2000). Linking school effectiveness and school improvement. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 206–231). New York: Falmer Press.

Schiefele, U., & Krapp, A. (1996). Topics of interest and free recall of expository text. *Learning and Individual Differences, 8*(2), 141–160.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*, 221–258.

Tamir, P. (1996). Science assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 93–129). Boston: Kluwer.

Tobias, S. (1994). Interest, prior knowledge and learning. *Review of Educational Research, 64*(1), 37–54.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators.* Washington, DC: The Falmer Press.