5 **Part Five** *of a Series of Five Reports*

High School Reform in Chicago Public Schools: A Snapshot of High School Instruction

August 2009

Joy K. Lesnick
Lauren Sartain
Susan E. Sporte
Sara Ray Stoelinga

SRI International®

These reports were produced by a team of researchers from SRI International and The Consortium on Chicago School Research. The research team included Daniel C. Humphrey, Marjorie E. Wechsler, Viki M. Young, Ashley Campbell, Patrick M. Shields, Maria I. Abasi, Lauren Cassidy, Raymond McGhee, Jr., and Samantha Murray from SRI; and Sue E. Sporte, Macarena Correa, Holly M. Hart, Joy K. Lesnick, Lauren Sartain, Sara Ray Stoelinga, Julia Gwynne, Jimmy Sebastian, and W. David Stevens from the CCSR.

The authors wish to acknowledge our partnership with the Chicago Public Schools in conducting this research, and particularly thank Bret Feranchak for his extensive assistance. In addition, we wish to thank Steve Cantrell and Ramona Thomas of the Bill & Melinda Gates Foundation for their guidance.

The research was supported by the Bill & Melinda Gates Foundation. The views, findings, and conclusions expressed herein are those of the authors and do not necessarily express the views of the foundation. Any formal endorsement by CCSR's Steering Committee, their organizations, or CCSR's directors should not be assumed.

High School Reform in Chicago: A Snapshot of High School Instruction

Summary of Findings

This is one of five pieces that comprise the evaluation report focused on high school reform in Chicago Public Schools. The Overview report summarizes findings from the three initiatives (AMPS, IDS, and Renaissance 2010) and presents cross-initiative findings. The remaining three reports are specific to each reform initiative.

In this brief, we report on the results of observations conducted across all three initiatives in Fall 2008 – a total of 78 classrooms in 17 high schools. For every classroom visited, a member of the research team used observational evidence to assign a rating of unsatisfactory, basic, proficient, or distinguished to each of 24 elements identified by Charlotte Danielson in her *Framework for Teaching* (2007). For this report, three focus elements were selected to illustrate larger constructs in our exploration of classroom practices:

Focus Element	Larger Construct
Teacher interaction with students	Classroom climate
Management of transitions	Classroom management
Quality of questions	Instructional practice

The elements, descriptions of the rating procedure, and qualitative examples of each level appear in detail in the full report, which follows this summary. The four primary findings described in the report are as follows:

1. **Among these elements, teachers we observed received the highest ratings for teacher interaction with students, followed by management of transitions, and then quality of questions.**

Seventy-two percent of teachers we observed received proficient or distinguished ratings on teacher interaction with students. Fifty-nine percent received a proficient rating or above for management of transitions, while only 30% received a proficient rating or above for quality of questions.

2. **Ratings varied across teacher experience, subject area, and grade level taught. In general, experienced teachers received higher ratings than new teachers, math and English teachers received higher ratings than science teachers, and 10th-12th grade teachers received higher ratings than 9th grade teachers. Although experienced teachers received higher ratings than new teachers, ratings for experienced teachers were still low.**

Teachers in their first or second year of teaching were observed in 21 of 78 (27%) of the classrooms. The data show that a smaller percentage of new teachers than experienced teachers achieved a rating of proficient or above on all three of the focus elements. In addition many teachers in at least their third year of teaching received ratings of unsatisfactory or basic: 25% of

experienced teachers on the teacher-student interactions, 44% on the management of transitions, and 64% on quality of questions.

Of the 78 classrooms observed, 25 were English classes, 33 were math classes, and 20 were science classes. Science classes received the lowest percentage of proficient ratings for management of transitions and quality of questions. In teacher-student interactions, however, ratings were generally similar across subjects.

Fifty of the 78 observations took place in 9th grade classrooms, 19 took place in 10th grade classrooms, and 9 took place in 11th or 12th grade classrooms. On average, we found that 9th grade classrooms received lower ratings as compared to 10th, 11th, and 12th grade classrooms in management of transitions and quality of questions. Ratings in student-teacher interactions were generally similar across grade levels.

3. Ratings of classrooms vary both between and within schools. Two "low" schools received low ratings across teachers, while three "strong" schools received high ratings across teachers. Twelve "mixed" schools had a range of teacher ratings. We found proficient teaching in schools regardless of incoming student ability.

Three of the 17 schools in our sample received consistently high ratings across the classes we observed, while 2 of 17 schools received generally lower ratings. The remaining 12 schools received mixed ratings across the classrooms we observed. In these schools, some teachers received higher ratings while others received lower ratings. While we observed highly rated classrooms where average student performance was high, we also observed highly rated classrooms in schools where average student performance was low.

4. Classroom climate and management are related to proficient levels of instruction. In classrooms where the climate rating was low, the classroom management rating was also low. In classrooms where classroom management was rated low, the quality of questions observed was also low.

The data show a strong relationship between classroom climate and classroom management. Of the 17 teachers who received a rating of unsatisfactory or basic on teacher-student interactions, only three received a proficient rating on managing transitions.

Similarly, the data show a strong relationship between classroom management and instructional practice. Of the 25 teachers who received a rating of unsatisfactory or basic on management of transitions, only four received a proficient rating on quality of questions.

High School Reform in Chicago: A Snapshot of High School Instruction

This is one of five briefs that comprise the evaluation report focused on high school reform in Chicago Public Schools. The Overview report summarizes findings from the three initiatives (AMPS, IDS, and Renaissance 2010) and presents cross-initiative findings. The remaining three reports are specific to each reform initiative. Improving teaching and learning is at the heart of every educational reform. Therefore, exploring what goes on in classrooms is an important source of information about reform implementation. This report focuses on classroom instruction across all three initiatives. Due to limitations of time and resources, we were only able to observe each classroom in the sample for one 45 minute class period; we call our data a "snapshot of instruction" at one point in time for just this reason.

We conducted observations across all three initiatives in Fall 2008 – a total of 78 classrooms in 17 high schools. After describing our sample and analyses, we provide an overview of the observations in these CPS classrooms using three illustrative elements chosen from Charlotte Danielson's *Framework for Teaching*. We then investigate variations in ratings across subjects, grades, teacher experience, and schools. Finally, we explore the relationship between elements to glean insights about areas where increased focus could support the goal of continued improvement.

Data Collection and Analysis

For the qualitative part of our evaluation, we used a stratified random sampling design to select schools in each initiative at each implementation wave (year 1, year 2, or year 3). We randomly selected five of 21 AMPS high schools, 13 of 43 IDS high schools, and 9 of 27 Renaissance 2010 high schools for a total of 27 high schools.

Teams of two or three researchers made a one-day intensive site visit to each of the 27 schools during October-November of 2008, interviewing principals, guidance counselors, and teachers of English, science, and math. The research team also observed classrooms in different grade levels and subject areas in the 17 schools that had been implementing the reform strategy for at least one year. The school contact (often a principal, vice principal, or front-office administrator) created the observation and interview schedule for the team. In all, we conducted a total of 78 classroom observations in 17 schools with an average of 5 observations per school.[1]

Researchers used Charlotte Danielson's *Framework*[2] as the classroom observation instrument. A Danielson *Framework* expert trained 13 researchers in how to use the rubric for rating teachers on 24 different elements in two of Danielson's four "domains" – classroom environment and instruction. Researchers visited classrooms for one class period (typically 45 minutes), recorded qualitative evidence, and made ratings for each element on a four-point scale: unsatisfactory, basic, proficient, or distinguished.

---

[1] In addition to the 78 observations and interviews, we conducted 90 other interviews in those same schools for a total of 168 interviews. We conducted 266 interviews in the full sample of 27 schools.

[2] Danielson, C. (2007). *Enhancing professional practice: A framework for teaching.* 2nd ed. Alexandria, VA: ASCD.

It is important to note that we approached this research on classroom practice with caution and caveats. We recognized that a variety of contextual factors influence the success or challenge of classroom instruction. It would be impossible for us to capture, measure, or even understand the variety of factors that are present during a one-time classroom visit. Therefore, we view our data collection pilot efforts as a "snapshot" of high school instruction during one class period at one point during the school year. Because of the stratified sampling approach, we believe that the classrooms we observed are generally representative of high school instruction across the reform initiatives that are currently being implemented in Chicago high schools, although it is possible that our in-school contact arranged our schedules to see only the "better" teachers.

The Classroom Observation Tool: Danielson's *Framework for Teaching*

Classroom observations in this research study were conducted using Charlotte Danielson's *Framework for Teaching*.[3] As described in depth in Danielson's *Enhancing Professional Practice: A Framework for Teaching* (2007), the framework identifies aspects of teachers' work that promote student learning based on research. It divides the complex activity of teaching into four domains: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. These domains are broken down into a set of 22 components that are then subdivided into 76 smaller elements. The framework defines four levels of performance – Unsatisfactory, Basic, Proficient, and Distinguished – for each element, providing a roadmap for the actions teachers can take to improve their instruction and, in turn, student learning.

We selected two domains for use in our classroom observations – The Classroom Environment (Domain 2) and Instruction (Domain 3). Although Planning and Preparation (Domain 1) and Professional Responsibilities (Domain 4) are important aspects of a teacher's professional practice, Danielson describes these domains as unobservable in the classroom. Therefore, in our classroom observations, we focused our observations and evidence on 10 elements in five components in Domain 2, and 14 elements in five components in Domain 3. The components and elements appear in Exhibit 1 below.

We selected three elements to portray our analyses and findings, which are underlined in Exhibit 1. Based on extensive discussion among the research team and a comparison of the ratings from other elements in the larger construct we believe that the ratings teachers receive on the following three elements can serve to illustrate three larger constructs of interest:

- 2.a.1: Teacher-interaction with students → Classroom climate
- 2.c.1: Management of transitions → Classroom management
- 3.b.1: Quality of questions →Instructional practice

[3] CPS has recently implemented a pilot project in 43 elementary schools using the Danielson *Framework* for teacher evaluations. Unlike our approach of using the rubric for a one time rating, the district will be following Danielson's guidelines to use the tool in a formative way as well as to provide evidence for teacher's summative efficiency ratings. Despite this difference, using the same instrument for observation affords us a similar language for discussion and was part of the motivation to use the Framework in this research study.

In the section that follows, we provide a description of each of the three focus elements we selected and what they represent.

Exhibit 1. Components and Elements Used in this Research Study[4]

Domain 2: The Classroom Environment	Domain 3: Instruction
a. Creating Environment of Respect & Rapport 1. Teacher Interaction with Students 2. Student Interactions with One Another **b. Establishing a Culture for Learning** 1. Importance of the Content 2. Expectations for Learning and Achievement **c. Managing Classroom Procedures** 1. Management of Transitions 2. Management of Materials And Supplies **d. Managing Student Behavior** 1. Monitoring of Student Behavior 2. Response to Student Misbehavior **e. Organizing Physical Space** 1. Safety and Accessibility 2. Arrangement of Furniture and Use of Physical Resources.	**a. Communicating with Students** 1. Expectations for Learning 2. Directions and Procedures 3. Explanations of Content 4. Use of Oral and Written Language **b. Using Questioning & Discussion Techniques** 1. Quality of Questions 2. Student Participation **c. Engaging Students in Learning** 1. Activities and Assignments 2. Grouping of Students 3. Instructional Materials and Resources 4. Structure and Pacing **d. Using Assessment in Instruction** 1. Monitoring of Student Learning 2. Feedback to Students **e. Demonstrating Flexibility & Responsiveness** 1. Response to Students 2. Persistence

Focus Element #1: Teacher Interaction with Students

Teaching depends on the quality of relationships between teachers and students. If teachers do not demonstrate that they respect students and vice versa, then the tasks of teaching and learning become more difficult. In the analyses presented here, we use this element to illustrate classroom climate. Exhibit 2 presents the way in which the *Framework* defines the four levels of performance for this element.

[4] Danielson, C. (2007). *Enhancing professional practice: A framework for teaching.* 2nd ed. Alexandria, VA: ASCD, pp. 3-4.

Exhibit 2. Levels of Performance for "Teacher Interaction with Students" Element[5]

	Teacher Interaction with Students
Unsatisfactory	Teacher interaction with at least some students is negative, demeaning, sarcastic, or inappropriate to the age or culture of the students. Students exhibit disrespect for the teacher.
Basic	Teacher-student interactions are generally appropriate but may reflect occasional inconsistencies, favoritism, or disregard for students' cultures. Students exhibit only minimal respect for the teacher.
Proficient	Teacher-student interactions are friendly and demonstrate general caring and respect. Such interactions are appropriate to the age and cultures of the students. Students exhibit respect for the teacher.
Distinguished	Teacher interactions with students reflect genuine respect and caring for individuals as well as groups of students. Students appear to trust the teacher with sensitive information.

Observational evidence of teacher-student interactions

In classrooms we observed, unsatisfactory and basic ratings were characterized by a marked lack of respect both from teacher to student and from student to teacher. In some cases, the lack of respect occurred when teacher had low expectations of students. For example, when one teacher said to the class, "We've done this before, right?", a student replied, "In, like, first grade." In other cases, the teacher's responses to student misbehavior were inconsistently applied. For instance, in one classroom a teacher repeatedly told students who were chatting and off task to "shut up" and "stop talking now," but students who were not participating in group work were not reprimanded.

On the other hand, in classrooms with ratings of proficient or distinguished, teachers acknowledged and validated student ideas and sources of frustration. *"I don't want you to think I don't appreciate the question because I do. It shows me you are thinking,"* one teacher explained to a student after pushing the student on a point. *"This stuff is frustrating,"* a student stated to a teacher. *"You need practicing, just like playing the drums,"* the teacher replied.

In addition, classrooms with positive ratings on teacher-student interactions were characterized by teachers establishing individual relationships with students based on their interests and identities. For instance one researcher described the following in an observed classroom:

> *Teacher has an individual relationship with students in the classroom setting-- calling students by name, checking in on how things are going-as well as in between periods, where a student spent time talking with the teacher about a sensitive issue going on at home.*

[5] Ibid, p. 66

Researchers observed distinct differences in the style of successful teachers in developing rapport. Some teachers used a nurturing approach, as described by a researcher in this classroom observation:

> *The teacher refers to most students as "baby" but not in a condescending manner; instead it seems to be a caring or motherly manner. The teacher mentioned that she liked what she was seeing to the whole class. 'I see some great stuff.'*

Other teachers relied upon a more "businesslike" relationship where academic expectations were clear and high. An example of this type of approach is illustrated by a researcher's observation in a classroom:

> *The teacher had a patient but assertive and serious tone when talking to students. (S)he called students by name and offered encouragement, as well as indicating concern for students' future. (S)he told students: 'Arm yourself for education [meaning be ready and get out of this what you need].' 'This may be the only chance you have to get this.' Students who gave wrong answers were not made to feel stupid, but (s)he also did not tolerate poor work.*

In both cases, teachers set the tone for respectful interactions between the teacher and students.

Focus Element #2: Management of Transitions

Managing the activities of a large number of students is a challenging task, but a smoothly functioning classroom is necessary for good instruction. Therefore, teachers must develop and implement procedures and routines that ensure a smoothly operating classroom and efficient use of time before they can focus on instructional tasks or goals. When transitions are successful, directions are clear and students know what to do. Little time is lost and the momentum of learning activities is maintained. Exhibit 3 presents the *Framework* rubric for this element.

Exhibit 3. Levels of Performance for "Management of Transitions" Element[6]

	Management of Transitions
Unsatisfactory	Transitions are chaotic, with much time lost between activities or lesson segments.
Basic	Only some transitions are efficient, resulting in some loss of instructional time.
Proficient	Transitions occur smoothly, with little loss of instructional time.
Distinguished	Transitions are seamless, with students assuming responsibility in ensuring their efficient operation.

[6] Ibid, p. 72.

Observational evidence of management of transitions

Classrooms characterized by a lack of routines and expectations about student behavior during transitions received low ratings. Students did not follow directions, and teachers lost instructional time trying to refocus students after a transition. For instance, one researcher noted:

> *Students were to count off into groups. They tried to shuffle around in the line, and they had to start the count over. Groups are assigned different areas of the room to sit. There is a lot of complaining and shuffling, and students are still trying to move into groups of their choice.*

Additional problems occurred when students did not play a role in routine tasks such as taking attendance, passing out papers, collecting assignments, or getting supplies and materials. As a result of this lack of student responsibility, we observed misbehavior and/or lost instructional time. One summary of a classroom observation stated:

> *Time was lost for late students, missing materials, taking attendance, and passing out paper and scissors.*

In the majority of classrooms we observed, a "bell ringer" activity was assigned at the beginning of class to focus students on an activity once they sat down during the new class period. In some classrooms with low "management of transitions" ratings, students did not complete these bell ringer activities in the allotted time, and then they complained about not finishing or being behind when the teacher went to the next activity. For example, one researcher wrote:

> *Students came in and the bell ringer was on the board. The first 15 minutes of the class were lost as the students laughed and fooled around, the vast majority never starting the bell ringer. The teacher announced that time was up for the bell ringer, and students protested, stating, 'You didn't give us enough time!'*

On the other hand, in classrooms we observed that were rated proficient or distinguished in managing transitions, teachers completed routine tasks while students were working. For example, teachers completed administrative activities during individual work time, small group time, or a bell ringer. One researcher wrote: *No instructional time was lost as the teacher took attendance as students worked on the bell ringer activity.*

Focus Element #3: Quality of Questions[7]

The questions that teachers ask help students explore new concepts, articulate their understanding, and promote deeper engagement in the material. High-quality questions promote student thinking, encourage them to make connections to previous knowledge or new ideas, and

[7] It is harder to get a proficient or distinguished rating on this element than on the other two focus elements. Data from the distribution of ratings across all 24 elements show that the majority of teachers we observed received low ratings on this element. It may be because high-level questioning techniques are inherently more difficult than other instructional tasks, or because this is an area that requires further attention for improvement, or both.

motivate them to consider new possibilities. Exhibit 4 presents the way in which the *Framework* defines the four levels of this element.

Exhibit 4. Levels of Performance for "Quality of Questions" Element[8]

	Quality of Questions
Unsatisfactory	Teacher's questions are virtually all of poor quality, with low cognitive challenge and single correct responses, and they are asked in rapid succession.
Basic	Teacher's questions are a combination of low and high quality, posed in rapid succession. Only some invite a thoughtful response.
Proficient	Most of the teacher's questions are of high quality. Adequate time is provided for students to respond.
Distinguished	Teacher's questions are of uniformly high quality, with adequate time for students to respond. Students formulate many questions.

Observational evidence of quality of questions

In classrooms we observed, teachers who asked questions without wait time or who supplied answers themselves received unsatisfactory or basic ratings on the quality of questions element. For instance, one researcher described the following in a classroom:

> *Teacher asks for what pattern students are seeing, but doesn't wait for an answer. Instead (s)he puts up a completed graph and shows them the answer. 'It's shown right here.'*

An additional common theme in classrooms rated unsatisfactory or basic in this element was that questions were intended to elicit specific (narrow) responses, and there was little follow-up questioning. For instance, a classroom observer noted:

> *The teacher's questions all require single responses and little critical thinking. Here is an example of an exchange between the teacher and students:*
> *Teacher: What is homo erectus? Who was walking on two feet?*
> *Students call out: Lucy!*
> *Teacher: Lucy. Wonderful. What does erectus mean?*
> *Students: Walking on two feet.*
> *Teacher: How big was her skull?*
> *Students: A softball.*

The kinds of questions asked in higher rated classrooms differed from those in lower rated classrooms. Students were asked questions that promoted reflection, reasoning, and dialogues such as "what if" and open-ended questions. We observed teachers asking students follow-up questions to promote deeper thinking and pushing students to support their position. For instance, teachers asked *"How do you know? Can you give me an example?"* The following description by a researcher comes from a classroom rated proficient in this element:

[8] Danielson, C. (2007). p. 82.

Many questions were factual to check student understanding of complex ideas and vocabulary. But scenarios were open-ended and thought-provoking with real-world application. Scenarios did have a "right" answer but more than just one.

Using questions allowed the teacher and students to reason through a process together and to move students through solving a problem. For example, in guiding students through a mathematical proof, one teacher stated:

"Let's take it one step at a time. Let's think about how we would do that. Just think. I want everybody to think. Let the wheels turn for a second. [Pause]. What's the next obvious thing to do? Sarah?"

Above are descriptions of our three illustrative elements. We next present a general description of the ratings observed by the research team followed by the four primary findings that emerged from this study

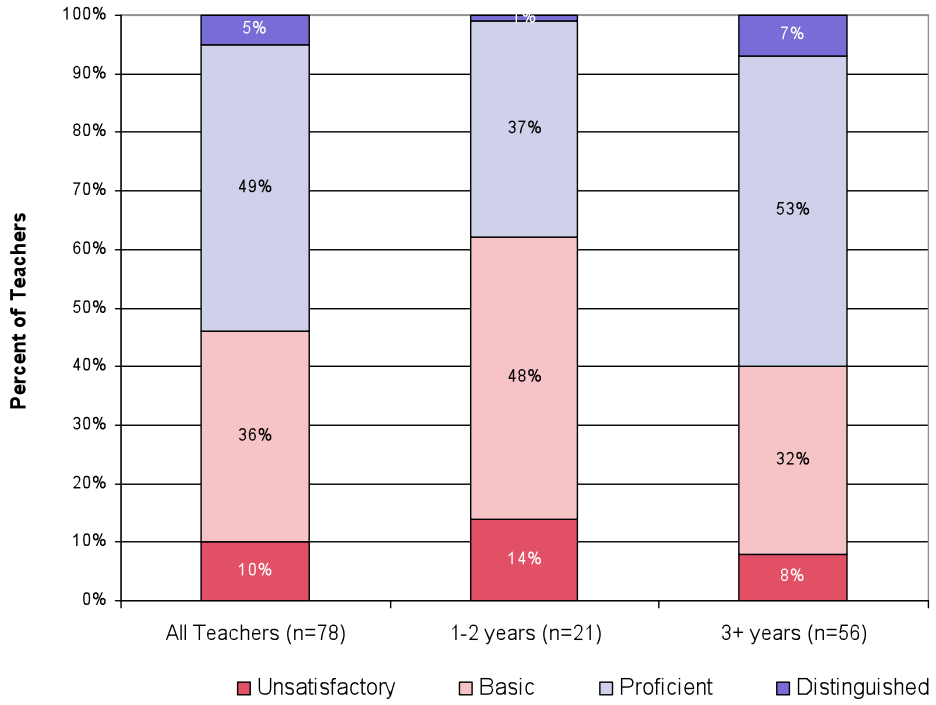
Findings and Discussion[9]

Ratings were collected for 24 elements in 78 classrooms for a total of 1872 possible ratings. If the observer did not have enough evidence to make a rating for one or more of the 24 elements, no rating was assigned. As a result, a total of 1598 ratings were recorded (85% of the possible total). Of these ratings, 10% were unsatisfactory, 36% were basic, 49% were proficient, and 5% were distinguished. Exhibit 5 displays the distribution of ratings across the observations of all teachers (first vertical bar), new teachers -- defined as teachers in their first or second year of teaching (second vertical bar), and experienced teachers -- those with 3 or more years of teaching experience (third vertical bar).[10]

[9] See Appendix A for notes on how to read the graphs in this section.

[10] 78 total teachers were observed. 21 were identified as new teachers, 56 were identified as having three or more years of experience, and level of experience for one teacher was not available in the data.

Exhibit 5. Distribution of all ratings for all, new, and experienced teachers

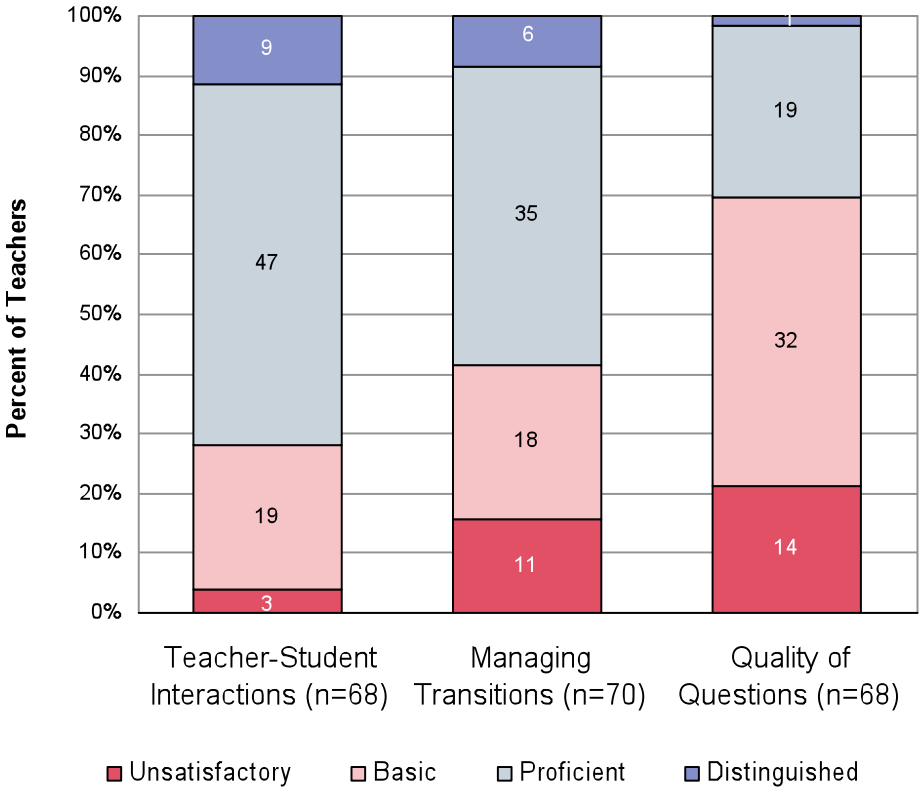


A full 60% of the ratings assigned to classrooms with experienced teachers were either proficient or distinguished (the top two sections of the vertical bar on the far right) as compared to 38% of the ratings for new teachers (middle vertical bar). Although it is not surprising that experienced teachers received higher ratings, 40% of all ratings for experienced teachers did not meet the proficient level, and is cause for further investigation. In the four specific findings presented below, we examine the variation of teacher ratings using the three focus elements of teacher-student interactions, management of transitions, and quality of questions described above.

Finding #1: **Among our 3 focus elements, teachers received the highest ratings for teacher interaction with students, followed by management of transitions, and then quality of questions. On average, teachers received low ratings for the quality of questions asked of students.**

Focusing on our three selected elements, we found higher levels of proficiency for teacher interaction with students as compared to managing transitions and quality of questions. Exhibit 6 displays the distribution of rubric ratings across all observed classrooms.

Exhibit 6. Positive ratings for teacher-student interactions are more common than positive ratings for managing transitions or quality of questions



In the first vertical bar, the distribution of ratings for the element "teacher interaction with students" is displayed. We use this element as an illustration of classroom environment. Note that in 28% of cases, teachers received ratings of unsatisfactory or basic (the percentage corresponding to the top of the second part the bar). Therefore, fully 72% (100% - 28%) of them received ratings of proficient or distinguished (top two parts of the bar). In the second vertical bar, 41% of classrooms received unsatisfactory or basic ratings, while the percent of proficient and distinguished ratings for "management of transitions" decreases to 59%. Finally, in the third vertical bar, we see that only 30% of classrooms received ratings of proficient or distinguished for the "quality of questions" element, and 70% of classrooms were rated as unsatisfactory or basic on this element.

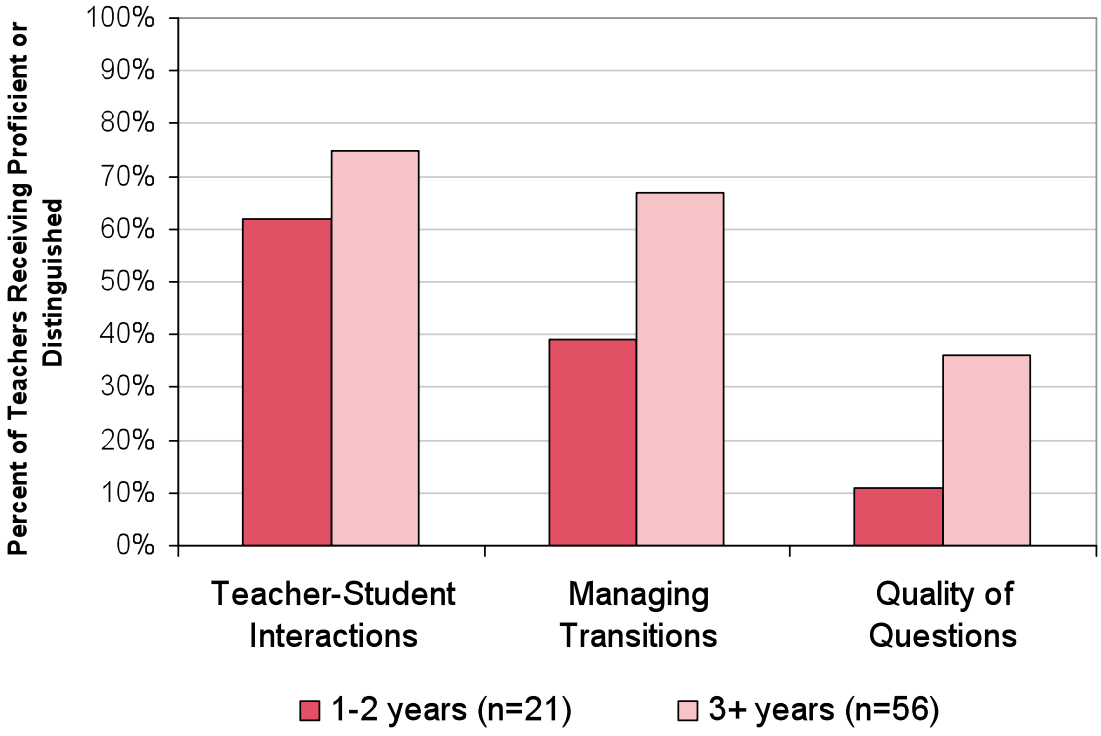
Based on our data, establishing a classroom climate with high levels of positive teacher-student interactions is the easiest of the three focus elements to achieve. The next hardest task is to manage transitions. Finally, asking high level questions is the hardest task of the three we describe here. Most classrooms did not receive a proficient rating on this element.

Finding #2: **Ratings varied across teacher experience, subject area, and grade level. In general, new teachers received lower ratings than experienced teachers, but ratings for experienced teachers are still low. Math and English teachers received higher ratings than science teachers, and 10th-12th grade teachers received higher ratings than 9th grade teachers.**

Variation in Ratings by Teacher Experience

Teachers in their first or second year of teaching were observed in 21 of 78 (27%) of the classrooms. The data show that a smaller percentage of new teachers achieved a rating or proficient or above on all three of the focus elements. In terms of teacher interaction with students, 62% of teachers in their first or second year of teaching ("new") received ratings of proficient or above, compared to 75% of their more experienced peers (teachers in at least their third year of teaching). The distributions are displayed in Exhibit 7.

Exhibit 7. Experienced teachers were rated higher on all 3 focus elements

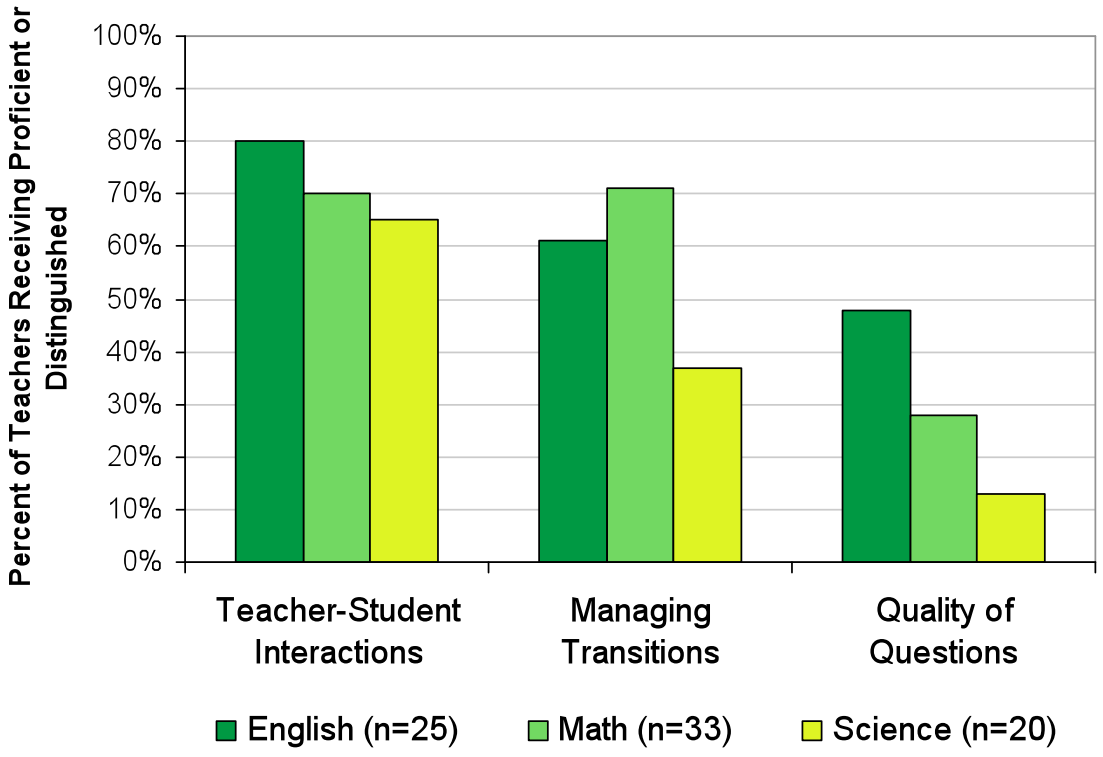


In terms of managing time through transitions between activities, 39% of new teachers received proficient ratings as compared to 56% of experienced teachers, and in the area of asking quality of questions, where we know that a majority of the teachers had difficulty, new teachers had more trouble on average when compared to their more experienced peers: 11% of them were rated as proficient, compared to 36% of those with more experience.

Variation in Ratings by Subject Area

Of the 78 classrooms observed, 25 were English classes, 33 were math classes, and 20 were science classes. In terms of our classroom environment rating, Exhibit 8 shows that 80% of English classes, 70% of math classes, and 65% of science classes were rated proficient or distinguished.

Exhibit 8. Science classes received lower ratings on all three elements.



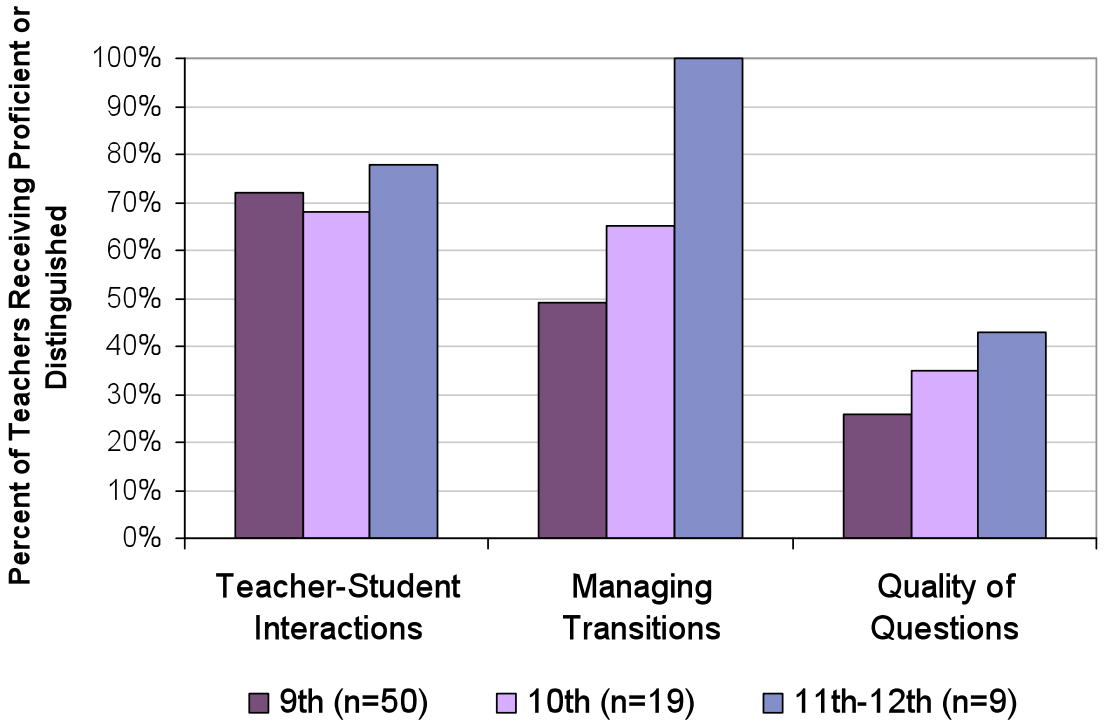
The ratings for management of transitions across subject area were more varied. 61% of English, 71% of math, but only 37% of science classes received proficient or distinguished ratings on the "management of transitions" element. This indicates that more instructional time is lost in science classrooms moving from one activity to another.

The ratings for quality of questions across also show that science classrooms had lower ratings than math or English classrooms. Only 13% of science classes received proficient ratings, compared to 28% of math classes, and 48% of English classes.

Variation in Rating by Grade Level

Of the 78 observations, 50 took place in 9th grade classrooms, 19 took place in 10th grade classrooms, and nine took place in 11th or 12th grade classrooms. On average, we found that 9th grade classrooms receive lower ratings as compared to 10th, 11th, and 12th grade classrooms in managing transitions and quality of questions, but not teacher-student interactions. Exhibit 9 shows that the majority of classrooms at all grade levels received ratings of proficient or above for teacher-student interactions.

Exhibit 9. Ninth grade teachers received lower ratings on managing transitions and quality of questions



Differences between grade are evident in management of transitions and quality of questions, however. Less than 50% of 9th grade classrooms received proficient or distinguished ratings on managing transitions, as compared to more than 75% of 10-12th grade classrooms. Similarly, 38% of 10th and 45% of 11th-12th grade classrooms were rated proficient or above on the quality of questions element as compared to 26% of 9th grade classrooms.

In sum, our data show that ratings on the three focus elements varied by teacher experience, subject, and grade level with new teachers, science teachers, 9th grade teachers faring worse than their peers. Because we know that new teachers received lower ratings than did experienced teachers across all 24 elements in our data, we examined whether new teachers were driving the low ratings reported for the three focus elements above. Upon further investigation, we found that new teachers are not clustered in one grade level or one subject area in our sample. Furthermore, it is not the case that all of the teachers that received low ratings

were new teachers. 40% of all ratings for experienced teachers were either unsatisfactory or basic.

Finding #3: **Ratings of classrooms vary both between and within schools. Two "low" schools received low ratings across teachers, while three "strong" schools received high ratings across teachers. Twelve "mixed" schools had a range of teacher ratings. Proficient levels of teaching were observed in some classrooms in schools with low average student performance.**

In only a few schools were ratings consistent across all of the teachers we observed – low or high ratings across all teachers in the school. More often, the ratings on the observation instrument varied between teachers within the same school, with highly rated teachers located in the same building as those with lower ratings. We turn next to a discussion of those results.

Based on teacher ratings on all 24 of the elements used for this study, we categorized schools in four ways: low, low-mixed, mid-range, and strong. Criteria used to place schools in these four categories are as follows:

- Low schools. More than 20% of ratings are unsatisfactory AND more than 70% of ratings were unsatisfactory/basic combined. Low schools are those where teaching was rated across the board as poor.
- Low-mixed schools. 1) More than 20% of ratings are unsatisfactory BUT less than 70% of ratings are unsatisfactory/basic combined; or 2) More than 70% of ratings are unsatisfactory/basic combined. At low-mixed schools, the quality of teaching was also generally low, but there were more proficient, and sometimes even distinguished, ratings at these schools. Most of the ratings given at low-mixed schools were at the basic level.
- Mid-range schools. Less than 70% of ratings are unsatisfactory/basic combined AND less than 70% of ratings are proficient/distinguished combined. Mid-range schools did not have a lot of unsatisfactory ratings, but they also did not have a lot of distinguished ratings.
- Strong schools. More than 70% of ratings are proficient/distinguished combined.

Exhibit 10 displays the differences among school categories on the ratings from our three focus elements: teacher-student interaction, managing transitions, and quality of questions. Across all three elements, low schools had the lowest ratings, followed by low-mixed schools, mid-range schools, and strong schools.

Exhibit 10. Schools we visited fell into four categories according to rubric ratings

Categories	*Initiative*	*Quartile ranking based on 8th grade test scores of current first time 9th graders*	*Teacher-student interactions (%U/B/P/D)*	*Transitions (%U/B/P/D)*	*Questions (%U/B/P/D)*
Low -2 schools -9 observations	2 IDS	1st quartile: 1 2nd quartile: 0 3rd quartile: 1 4th quartile: 0	U: 22% B: 22% P: 44% D: 11%	U: 75% B: 25% P: 0% D: 0%	U: 63% B: 38% P: 0% D: 0%
Low-mixed -5 schools -23 observations	3 IDS 2 Ren10	1st quartile: 3 2nd quartile: 1 3rd quartile: 1 4th quartile: 0	U: 4% B: 39% P: 52% D: 4%	U: 21% B: 37% P: 37% D: 5%	U: 22% B: 56% P: 22% D: 0%
Mid-range -7 schools -31 observations	1 AMPS 2 IDS 4 Ren10	1st quartile: 2 2nd quartile: 4 3rd quartile: 0 4th quartile: 1	U: 0% B: 26% P: 65% D: 10%	U: 4% B: 32% P: 61% D: 4%	U: 15% B: 52% P: 30% D: 4%
Strong -3 schools -15 observations	1 IDS 2 AMPS	1st quartile: 0 2nd quartile: 0 3rd quartile: 0 4th quartile: 3	U: 0% B: 0% P: 73% D: 27%	U: 0% B: 13% P: 60% D: 27%	U: 8% B: 38% P: 54% D: 0%

* 8th grade ISAT scores from spring 2008, math and reading scores combined. Achievement academies ,alternative schools and special education schools were removed before quartile rankings assigned.

It is important to note that for schools identified as having high quality instruction across the board, incoming student achievement in math and reading is high. Therefore, for schools in our strong category, it is difficult to disentangle the effect of having high achieving students in the classroom with a teacher's ability to deliver high quality instruction. This finding is not surprising. However, our data show no relationship between ratings on our focus elements and incoming achievement in the other three categories of schools: low, low-mixed, and mid-range. This evidence suggests that having under-performing students in the classroom does not necessarily mean that the instruction will receive low ratings on the observation instrument used in this study. Teachers can have students who are below grade level and still be effective teachers. Similarly, having higher performing students in the classroom does not ensure high quality instruction. For example, freshmen at one of the schools in the mid-range category had average incoming standardized test scores in the top 25% of all high schools in CPS. Because the majority of the schools we visited fell into low-mixed and mid-range categories, we focus our analysis on the instruction that we observed in those schools.

Within our sample of 78 teachers, 16 teachers received no unsatisfactory ratings and no more than one basic rating on any of the 24 elements we observed. Seven of these teachers worked in schools categorized as strong, and again we note that it is difficult to disentangle the relationship between high-achieving students and high quality instruction. However, more than

half of these teachers worked in other types of schools: six teachers taught in mid-range schools, two taught in low-mixed schools, and one taught in a low school.

Overall these teachers exhibited mastery of establishing a positive classroom climate, managing transitions, and utilizing engaging instructional techniques. These teachers work with students who are comparable to students in other classrooms and schools in terms of prior achievement. Yet these teachers have demonstrated proficient-level skills. They may be an untapped school-level resource for support and mentorship for their colleagues.

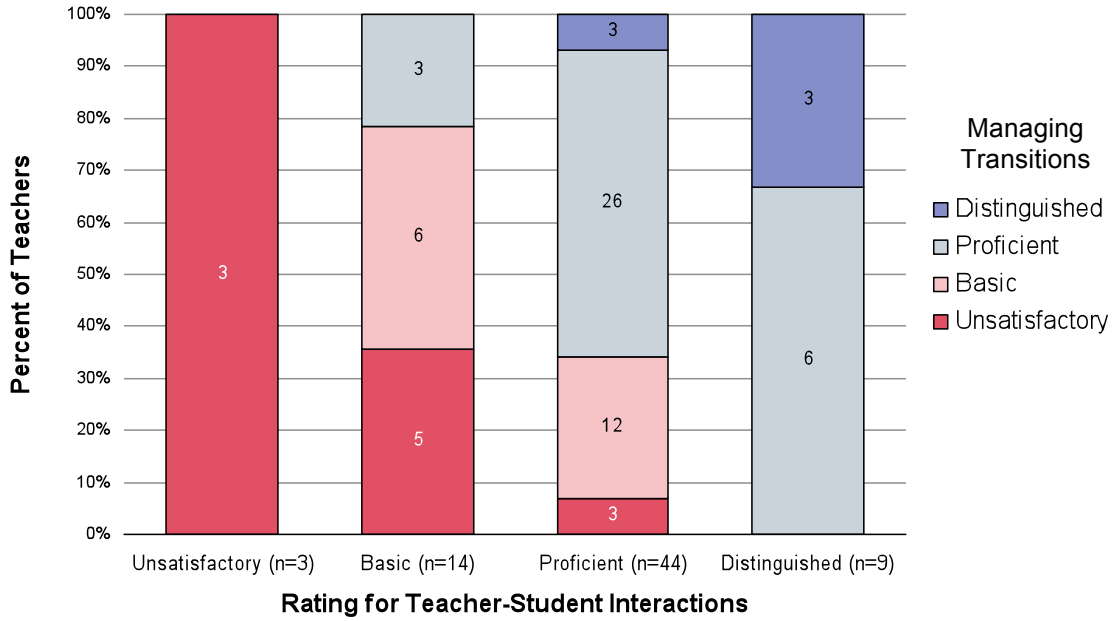
Finding #4: **In classrooms where the climate rating was low, the classroom management rating was also low. In classrooms where classroom management was rated low, the quality of questions observed was also low.**

Classroom Climate and Classroom Management

Our classroom observation data provide the opportunity to analyze the relationships between the ratings teachers received on different elements in the Danielson *Framework*. Given a particular rating on respect and rapport, do we see patterns in ratings the same teachers received in elements of classroom management? Analyzing ratings data with this question in mind, we do find a relationship. Exhibits 11-12 display the results of cross-tabulations of (1) student-teacher interactions and management of transitions, and (2) management of transitions and quality of questions and are interpreted slightly differently than the figures that appeared above.

In Exhibit 11, each vertical bar represents all of the teachers who received a particular rating for teacher-student interactions. The leftmost bar represents all those who received ratings of unsatisfactory, the next bar represents those who were rated basic, and so on. Within each of the 4 bars, each section represents the ratings that those same teachers received on the management of transitions elements. For example, starting with the first vertical bar, we can see that three teachers received ratings of unsatisfactory on teacher-student interactions. Of those, all three *also* received unsatisfactory ratings on management of transitions as noted by the dark color. In the second vertical bar, we see that 14 teachers (the sum of the numbers in the bar) received a rating of basic on the teacher-student interaction element. Of those 14, five received a rating of unsatisfactory on management of transitions, (the darkest part of the bar), six received a rating of basic (the next part of the bar), and three received a rating of proficient (the third part of the bar). In sum, of the 17 teachers in the two bars who received a rating of unsatisfactory (n=3) or basic (n=14) on teacher-student interactions, only three received a proficient rating on managing transitions.

Exhibit 11. Teachers rated low on teacher-student interactions also received low ratings for managing transitions

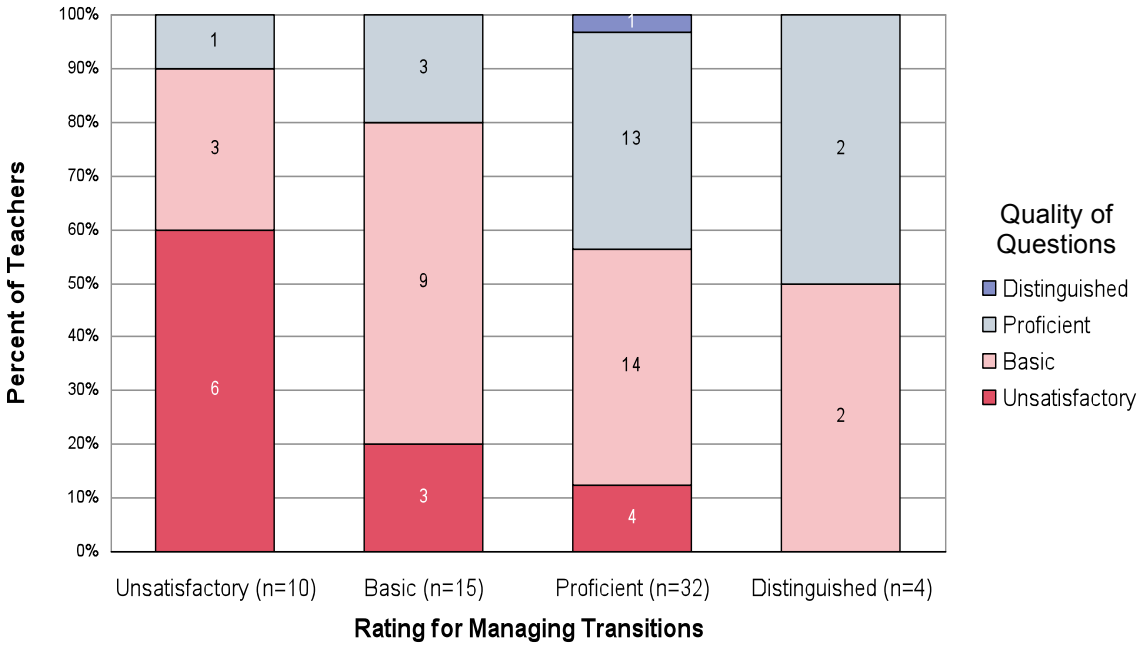


Data from this "snapshot" of instruction show a clear relationship between ratings for teacher-student interactions and managing transitions. Teachers who were able to create a foundation of positive teacher-student interactions were more likely to have higher ratings in classroom management. In turn, it is also likely that strong management practices reciprocally positively influence classroom climate. It is also clear, however, that establishing rapport with students did not necessarily lead to strong classroom management. Of the 53 teachers who received a rating of proficient or distinguished on teacher-student interactions (the 2 bars on the right), 15 were rated as basic or below on managing transitions. In this respect, creating a positive classroom climate was necessary but not sufficient for successful classroom management.

Classroom Management and Instructional Demand

We next considered the relationship between classroom management and instructional demand, and found a strong relationship between these two aspects of teaching, as displayed in Exhibit 12. Of the 25 teachers who received a rating of unsatisfactory (first bar, bottom two sections, n=7) or basic (second bar, bottom two sections, n=18) on managing transitions, only four received a proficient rating on quality of questions (third sections in both bars). However, of the 36 teachers who received a proficient or distinguished rating on managing transitions, 20 were rated as basic or below for quality of questions.

Exhibit 12. Teachers rated low on managing transitions also received low ratings on quality of questions



There appears to be a relationship between classroom management and instructional practice. As in the case of the relationship between classroom climate and classroom management, positive classroom management is necessary but not sufficient for positive outcomes in instructional practice. And, as in the case of the relationship between climate and management, it is likely that there is a reciprocal relationship between strong instructional practice and classroom management; that challenging instruction practices in turn reinforce classroom management.

Using the Danielson *Framework*

Throughout the evaluation, we considered the observational research presented here as a pilot study with the intention of providing us with insights into various components of teacher practice and the relationship between these components. As in all observational research, this study has some methodological limitations. This section presents the limitations we experienced and lessons we learned for future research.

Limitations

First, the selection of exactly which classrooms we visited in each school was not completely random. We asked school coordinators, often the principal or assistant principal, to help us schedule a series of approximately five classroom observations across grade levels and subject areas within one day. We also asked them to schedule interviews with those five teachers plus interviews with approximately six to eight additional teachers and administrators, for a total of 16-18 "research events" to be covered in one day by a team of two or three researchers. The logistics of this scheduling perhaps mitigated against our seeing only the "best" teachers in the building, but it is also possible that the coordinator was able to avoid having us see the "worst" teachers in the building.

Second, we only visited each classroom once, and the day was of our choosing. It is possible that we visited a number of teachers at a point in their overall lesson plan that was not representative of what typically happened in each classroom. However, we visited only nine classrooms where the teachers were either testing all period or which were math computer lab days, where we were unable to view such elements as whole-class questioning techniques or to observe teacher-created activities and assignments.

Third, each classroom was visited by only one researcher, so there was no built-in inter-rater reliability in the ratings teachers received. The relatively large number of researchers relative to the number of observations may have guarded against persistent rating bias. Thirteen researchers participated in the observations, typically observing about six different classrooms across three or four schools. All were trained in the use of the rubric over four intensive half day sessions discussing the nuance of each element and practicing by rating video-taped lessons. At the end of the training sessions there were no serious cross-rater differences; and across the 78 observations we saw no systematic cases of researcher rating bias. However, it still could be the case that if two observers had rated the same teacher they would have had slightly different ratings on at least some of the elements.

Lessons Learned

Before we chose this observation rubric we engaged in a rather lengthy discussion of whether we should rate teachers on the smaller and more precisely defined "elements" or whether we should base our ratings on the larger and slightly more holistic "components." Each component typically has two to three elements. For example, component 3b, Using Questioning and Discussion Techniques, had three elements: quality of questions, discussion techniques, and student participation in discussion. We decided to give teachers a separate rating on each of the elements. While this may have helped us deconstruct what goes into "questioning and discussion techniques," using a multitude of individual elements made it difficult to effectively summarize our findings. For this brief, our goal was to present the results in a clear and understandable way. So we chose elements that seemed to convey the general sense of the larger construct. Considering all of the Framework's elements are important when trying to measure the a larger construct, but for the purposes here, we concluded that reporting separate ratings on each of 24 elements would not be as useful as focusing on three elements to illustrate the larger constructs.

Implications and Conclusions

Our goal in this report has been to provide a snapshot of instruction; to provide a sense of the overall look and feel of selected CPS high school classrooms during one class period in fall 2008. We have also aimed to provide a sense of the variation in quality of classroom climate, classroom management and instructional practice within and across schools and across subjects, grade levels and teacher experience. Finally, we considered the relationship between classroom climate, classroom management and instructional practice in order to gain a better understanding of the way they interact and influence classrooms. We believe they can help inform CPS in general and as well as each of the three initiatives about ways they can strengthen classroom instruction.

While teachers have generally positive interactions with their students, most are not rated above basic on the quality of questions they ask their classes.

We showed that, overall, approximately 70% of teachers received proficient or distinguished ratings in teacher-student interactions and that 60% of them received proficient or distinguished ratings in management of transitions. However, only 30% received these high ratings in quality of questions. While we only discuss these three elements in this brief, we note that the same general pattern emerges if we look at other elements measuring similar constructs. Clearly more work needs to be done in the area of providing all students with the opportunity to learn complex material that demands that they go beyond one-word answers and that they are presented with challenging assignments.

There is variation in ratings across teacher experience, subject areas, and grade level taught. Understanding this variation can help schools and initiatives provide targeted professional development.

We noted variation in ratings across teacher experience, subject areas, and grade level taught. First, we note that new teachers were rated lower on all three focus elements. While this may not be surprising, if our sample is representative of the distribution of teacher experience across CPS, then more than a quarter of classes are being taught by teachers that need more intensive support. Second, English classes were rated highest in terms of managing transitions and quality of questions, followed by math classes and then science classes. While this preliminary consideration is based on a relatively small number of observations in science classes, it is also clear that teachers in math and science classes may need additional supports in classroom management techniques and instructional practice, especially as indicated by quality of questions. Finally, we found systemically lower ratings for teachers in ninth grade across all three focus elements. Is it because of the unique needs of ninth grade students? Is it because ninth grade classrooms tend to be staffed by teachers with less experience or less ability to claim teaching assignments in upper grades? Is it a combination of these and other factors? Seeking the source of this variation is essential in targeting and improving instructional practices in ninth grade.

There is variation both between and within schools. School-wide factors probably play a role in both strong and weak teacher performance.

We found variation both between and within schools. At two schools in our sample, no classrooms received a rating higher than basic on managing transitions or on quality of questions, and more than 60% of classrooms in those schools received an unsatisfactory rating on these elements. At the other extreme, in three of our observed schools no classrooms received a rating of unsatisfactory in teacher-student interactions or in managing transitions, and more than half of all classrooms were above basic on all three of our indicators. Given our small sample size, it is possible that the classrooms we visited in these five schools were not typical of the classrooms in the whole school. Certainly these ratings don't occur in a vacuum—there must be other school level factors at work. However, they are NOT all related to the incoming achievement of the students in these schools, as we visited other schools with similarly under-prepared students which had overall better classrooms. And we also visited a school with students in the top quartile of preparation whose classrooms were not uniformly as good as others who had similar students.

We also saw variation in classroom ratings in 12 of our 17 schools. This variation was sometimes within *teacher* (that is, some teachers received high ratings on some of our 24 elements and low ratings on others), and sometimes within *school (*that is, some teachers in a school received generally strong ratings, while other teachers in the same school received generally weak ratings). Both types of variation can provide a strong foundation on which a school can build a good instructional program.

Strong classroom climate and classroom management are necessary but not sufficient to ensure good instruction.

We show empirically what seems like "common sense:" if teachers can not maintain mutually respectful interactions with their students they do not do well with managing classroom management routines. And, if teachers can't handle classroom management routines, they do not do well with such elements as asking quality questions.
While this may seem like an obvious finding, schools and initiatives may need to be reminded that professional development related only to the delivery of content may not be effective—there may need to be additional support in classroom management skills.

Appendix A

A Note on Reading the Bar Graphs

A number of graphs and tables are presented in this report to illustrate the distribution of unsatisfactory, basic, proficient, and distinguished ratings across the 78 classrooms that were observed.

When reading and interpreting the stacked bar charts, keep the following in mind:

1. Each vertical bar adds up to 100%. The percentage of unsatisfactory ratings appears at the bottom of the bar. Reading from bottom up, basic is placed on top of unsatisfactory. On top of that is the percentage of teachers receiving proficient ratings, with distinguished at the top of the bar.

2. The numbers in each of the sections of the vertical bars represent the actual number of classrooms in which that rating was observed. If a rating was recorded in all classrooms, then the numbers will total to 78. In some cases, the researcher did not have enough information to make a rating. Therefore, the total number of observations (and thus, the denominator for calculating the percent of classrooms) is less than 78.

3. To calculate the percent of teachers rated as unsatisfactory, read the percent on the vertical axis corresponding to the height of the darkest part of the bar; use a similar method to get the percent of teachers rated either unsatisfactory or basic. To calculate the percent of teachers rated proficient or above requires a little arithmetic. It's probably easiest to note the percent of teachers receiving basic or below and subtract that number from 100%.

When reading and interpreting the bar charts for variation between teacher experience, subject, and grade, the vertical bars represent the percent of teachers receiving proficient or distinguished ratings on the element. The number of teachers in each category is displayed across the horizontal axis.