THE QUALITY OF THE TEACHING

OF

MATHEMATICS, SCIENCE AND TECHNOLOGY IN K-12 CLASSROOMS IN NEW YORK STATE

A SUMMARY OF FINDINGS

Mark St. John
Becky Carroll
Jeanne Century
Costanza Eggers-Pierola
Judy Hirabayashi
Nina Houghton
Sybillyn Jennings
Felisa Tibbitts
Ruth Von Blum

April 1999

INVERNESS RESEARCH ASSOCIATES

THE QUALITY OF THE TEACHING OF MATHEMATICS, SCIENCE AND TECHNOLOGY IN K-12 CLASSROOMS IN NEW YORK STATE

A SUMMARY OF FINDINGS[1]

Introduction

For the past five years, Inverness Research Associates[2] has been the evaluator of the New York Statewide Systemic Initiative (NYSSI), a five-year, ten million dollar investment of the National Science Foundation (NSF) in the improvement of mathematics, science and technology (MST) education in New York State. During the first three years of this reform effort, we at Inverness Research focused most of our evaluation effort on documenting the progress of ten NYSSI pilot schools – also called "research and development" (R&D) schools.[3]

The later years of the NYSSI evaluation focused on strategies that were directed more broadly toward building the capacity of the state as a whole – for example, the development of the New York State MST Standards[4] and assessments. Not surprisingly, both state leaders and the NSF became interested in the progress of the state's MST standards-based reform effort. They were curious about the degree to which the new state standards were helping to improve the quality of instruction in the state's classrooms. Before they could begin to answer that question, they

---

[1] This summary of findings is derived from Part I of the report *New York State MST Landscape Study: A Study of the Current Status Of the Teaching of Mathematics, Science and Technology in New York State: The Progress of Standards-based Reform* which is available on the Inverness Research Website (www.inverness-research.org). We are also creating a separate summary of findings called: *Issues and Challenges for MST Standards-based Reform in New York State.*

[2] Inverness Research Associates, a small research and evaluation firm, studies educational change investments around the country and acts as evaluators for many NSF-funded systemic initiatives. The leadership of the New York State Systemic Initiative hired Inverness Research to be its internal evaluator, beginning in 1994.

[3] The lessons we learned in this phase of the evaluation resulted in a document called *The Principals of Reform: Supporting Mathematics and Science Teaching in Your School - A Handbook for Elementary and Middle School Principals.* This report is also available from the Inverness Research Website (www.inverness-research.org).

[4] *The Learning Standards for Mathematics, Science and Technology (MST).* The New York State Education Department, the University of the State of New York, Albany NY. March 1996.

faced an even simpler question: what was the current status and quality of the teaching in New York's MST classrooms?

In order to address this question, the Inverness Research team revised the evaluation plan of the NYSSI to include "The New York State Landscape Study." This effort included the careful study of seven randomly-selected school districts across the state, examining MST in each district by looking carefully at the actual instruction taking place in the districts' K-12 classrooms. Ultimately, we obtained a picture of the nature and quality of MST teaching by observing a total of 156 math, science and technology lessons across grade levels (K-12).[5]

The Study of MST in Seven New York School Districts

Selecting the Districts

In the fall of 1997, the New York State Education Department (SED) provided Inverness Research with a list of all state districts categorized by one of six district types.[6] The State Department of Education picked for us a stratified random sample of seven[7] districts in each of those categories. Within each grouping we asked that the districts selected cover the spectrum in terms of geographic location and size. Thus, our sample included large and small districts, rural and urban districts, and high-need and low-need districts.

---

[5] It is interesting to note that by comparison the TIMSS classroom studies involved observation of 100 US classrooms, 100 German classrooms, and 50 Japanese classrooms.

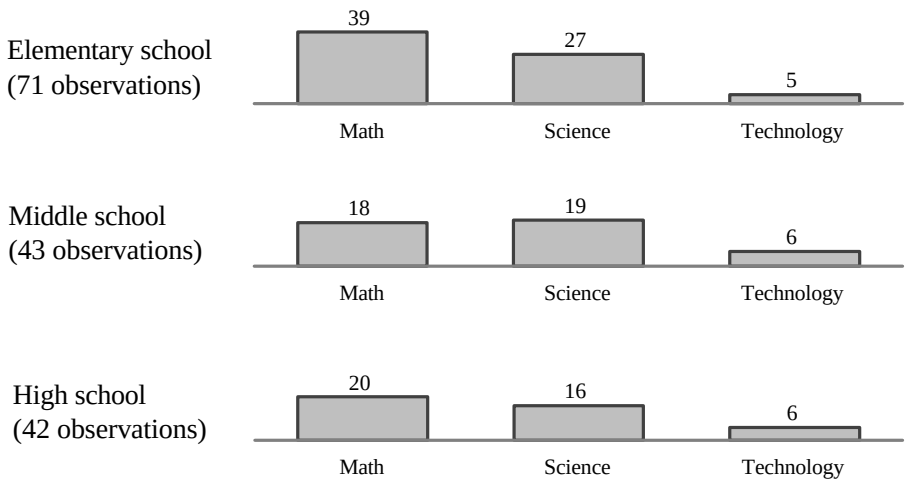
[6] New York State actually classifies its district by five main categories. These are New York City Public Schools; Large City Districts; Other City Districts; Suburban Districts; and Rural Districts. These classifications are based on the districts' geographical, political and employment characteristics. Within these categories the State Department of Education further grouped districts by their level of "need" so that six categories emerged. These included: 1) high-need, large city districts; 2) high-need, small city districts; 3) low-need, suburban districts; 4) average-need suburban districts, 5) average/low-need rural districts; and 6) high-need rural districts.

[7] In order to refine our methodology before visits in the other six districts occurred, all protocols were first used in a seventh "pilot" district.

The Classroom Observations

When a district agreed to participate, our researchers organized a three to five day visit to the district. A total of 156 classrooms were observed in this study. Our statewide sample included 77 math classes, 62 science classes and 17 technology lessons.[8] Almost half of the classrooms we observed were in elementary schools (46%); the rest of the observations were divided almost equally between middle school classrooms (28%) and high school classrooms (27%). Between 15 and 31 observations were conducted in each district (an average of 22 classrooms in each). With one exception, we observed lessons at the elementary, middle and high school levels in all of the districts we visited.

Types of classrooms observed (N = 156)



[8] It is important to note we saw both instructional technology and technology education lessons under the umbrella of "technology" classes. Seven of our observations were instructional technology classes and ten were technology education lessons.

The Horizon Protocol

In conducting the classroom observations that were a part of this study, researchers used an observation protocol developed by Horizon Research Institute, Inc. (HRI) to rate each classroom observation. The protocol was designed specifically for use in the NSF-funded Local Systemic Change (LSC) districts. It is important to note that the protocol is designed specifically to measure the extent to which classroom practice reflects the vision of MST instruction laid out in the national standards documents.[9] The protocol, in particular, emphasizes inquiry, standards-based content, and an equitable and student-centered classroom culture. Using this protocol researchers note the number of students, the classroom's resources and the focus of the lesson, and then rate the lesson's design, its implementation and content and the overall culture of the classroom. Researchers also make summary judgments about the lesson's likelihood of contributing to student understanding of and interest in the discipline, and an overall rating is given. These dimensions and the overall ratings system are described in more detail in the appendices of our full evaluation report.

The Horizon protocol gives high ratings only to those classrooms that are sound in the content they are teaching, that promote an inquiry-based approach, that value student thinking, that are inclusive and supportive of all students in the classroom, and that seek to make lessons relevant to today's culture and students' native interests. The rating scale ranges from 1 to 5 with 1 being the lowest and 5 being the highest. Level 1 observations are characterized as "ineffective instruction" - exemplified either by a predominance of "passive learning" or "activity for activity's sake"; Level 2 lessons show "elements of effective instruction"; Level 3 classrooms are taught by teachers whose practice is showing the "beginning stages of effective instruction" (and can be distinguished by a "low," "solid" or "high" rating); Level 4 lessons are "accomplished"; and Level 5 instruction is "exemplary."

---

[9] *National Council of Teachers of Mathematics (NCTM), Commission on Teaching Standards for School Mathematics* (1991). Professional Standards for teaching mathematics. Reston, VA; and *National Research Council (NRC), National Science Education Standards* (1996). Washington DC: National Academy Press.

From the HRI Protocol: Capsule Description of the Quality of the Lesson

Level 1: Ineffective Instruction

There is little or no evidence of student thinking or engagement with important ideas of science / mathematics. Instruction is *unlikely* to enhance students' understanding of the discipline or to develop their capacity to successfully "do" science / mathematics. Lesson was characterized by either:

Passive Learning: Instruction is pedantic and uninspiring. Students are passive recipients of information from the teacher or textbook; material is presented in a way that is inaccessible to many students.

Activity for Activity's Sake: Students are involved in hands-on activities or other individual or group work, but it appears to be activity for activity's sake. Lesson lacks a clear sense of purpose and / or a clear link to conceptual development.

Level 2: Elements of Effective Instruction

Instruction contains some elements of effective practice, but there are *substantial problems* in the design, implementation, content, and / or appropriateness for many students in the class. For example, the content may lack importance and / or appropriateness; instruction may not successfully address the difficulties that many students are experiencing; etc. Overall, the lesson is *quite limited* in its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully "do" science / mathematics.

Level 3: Beginning Stages of Effective Instruction

Instruction is purposeful and characterized by quite a few elements of effective practice. Students are, at times, engaged in meaningful work, but there are *some weaknesses* in the design, implementation, or content of instruction. For example, the teacher may short-circuit a planned exploration by telling students what they "should have found," instruction may not adequately address the needs of a number of students; or the classroom culture may limit the accessibility or effectiveness of the lesson. Overall, the lesson is *somewhat limited* in its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully "do" science / mathematics.

Level 4: Accomplished, Effective Instruction

Instruction is purposeful and engaging for most students. Students actively participate in meaningful work (e.g., investigations, teacher presentations, discussions with each other or teacher, reading). The lesson is well designed and the teacher implements it well, but adaptation of content or pedagogy in response to student needs and interests is limited. Instruction is *quite likely* to enhance most students' understanding of the discipline and to develop their capacity to successfully "do" science / mathematics.

Level 5: Exemplary Instruction

Instruction is purposeful and all students are highly engaged most or all of the time in meaningful work (e.g., investigations, teacher presentations, discussions with each other or teacher, reading). The lesson is well-designed and artfully implemented, with flexibility and responsiveness to student needs and interests. Instruction is *highly likely* to enhance most students' understanding of the discipline and to develop their capacity to successfully "do" science / mathematics.

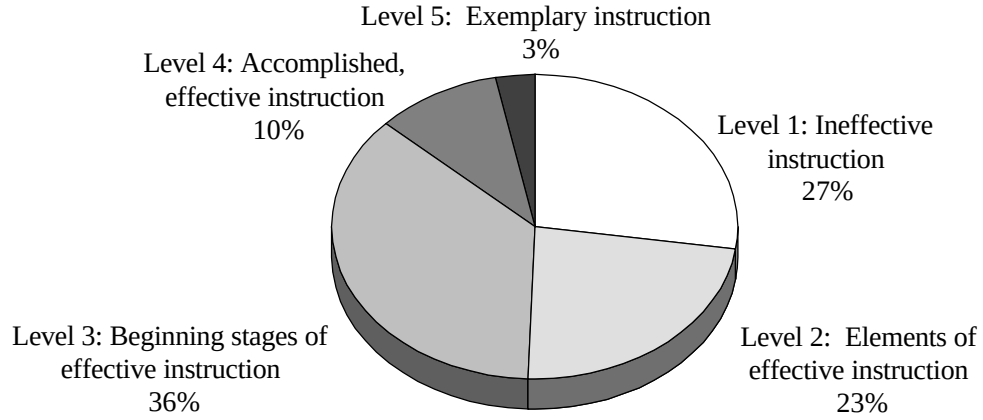
In what follows, we present a summary of our data that gives the reader an overall picture of the quality of MST teaching and learning in New York State. We have included data and observations on the overall quality of the lesson, the culture of the classrooms observed, and the likelihood of the lesson to enhance student ability and stimulate interest in the MST disciplines. We also have included data summaries that illustrate examinations of the differences between math, science and technology lessons, the comparison of the quality of elementary, middle, and high school lessons, and finally, differences in quality between different types of districts.

A Summary of Findings

The Quality of MST Teaching

Across the state, and within each district, there was a wide range in the quality of the MST instruction in the 156 lessons we observed. While we did see some strong lessons and exemplary teaching, we did not see many lessons reflecting the vision for classrooms that is laid out in the national standards documents. Less than one-fifth of the lessons we observed met these criteria. In fact, 50% of all lessons were classified as either Level 1 or Level 2 by researchers. Another third (36%) were rated as "3," the beginnings of effective instruction. Just over ten percent of all observed lessons (13%) were rated by our researchers as "accomplished or exemplary."

Overall quality of the MST lessons observed in New York



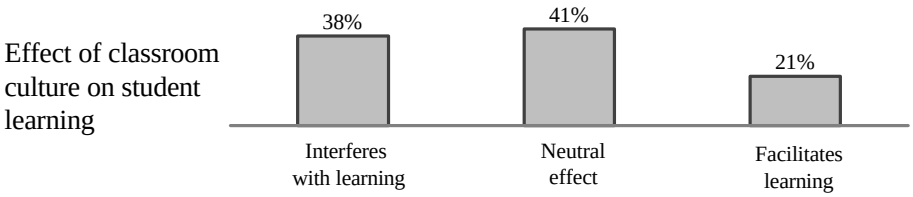
Culture of the Classroom

Researchers not only watched the details of the lessons that were taught, but they also examined the underlying culture of the classrooms they visited. Classroom culture within the Horizon protocol includes such elements as: the degree to which all students are actively engaged in the lesson; the extent to which students and the teacher work collaboratively; the extent to which the challenging of ideas and diverse approaches are valued and encouraged; and the degree to which the lesson provides and promotes equitable learning opportunities for all students. All of these elements of classroom culture can greatly contribute to or undermine the effectiveness of any given lesson for any particular student.

The results of our ratings of classroom culture are telling. In 21% of all the classrooms we visited we found a classroom culture that was healthy, exciting, and equitable, encouraging all students to become interested in and engage fully in the learning of the discipline. In 38% of the

classrooms that we visited, we saw classroom cultures that were deficient to the extent that they were actually interfering with students' learning.[10]

To what extent does the classroom culture facilitate learning?

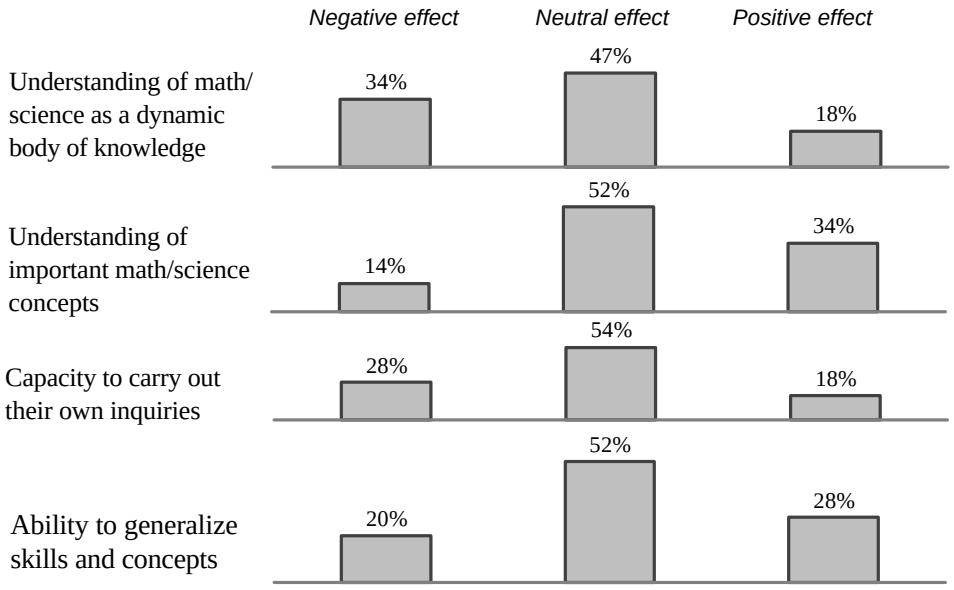


Enhancing Student Ability and Interest in the Discipline

Researchers also made judgements about the likely effects of each lesson on students' understanding of the important concepts and principles of math, science or technology. This evaluation was based on the lesson itself but also on pre- and post-interviews with teachers and students. Just over one-third (34%) of the lessons we observed were judged as likely to have a positive effect on students' understanding of "important math/science concepts." However, in only 18% of the lessons we observed did we rate the classroom as likely to enhance a students' capacity to carry out their own inquiries or increase their understanding of math/science as a dynamic body of knowledge.

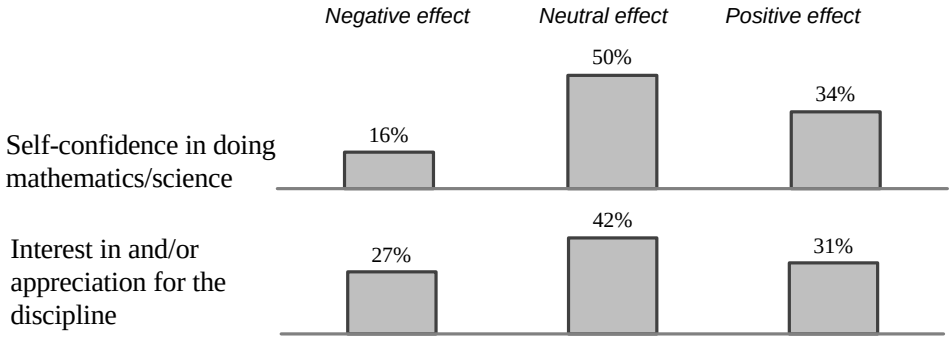
[10] Upon reviewing these results one administrator in the State Department of Education said, "These [classroom culture] results, I think, tell the whole story. What good is it if our classrooms are not places that encourage students to learn, and not places that are likely to interest them in further learning of math or science or technology…?"

Likely effects of the lesson on student understanding and skills



Thirty-four percent of lessons were characterized as positively affecting students' self-confidence. Moreover, 31% of lessons were seen as having a positive effect on students' interest in and appreciation for the discipline. These positive percentages notwithstanding, note should be taken of the majority of lessons we observed that were rated has having a neutral or negative effect.

Likely effects of the lesson on student self-confidence and interest in the subject

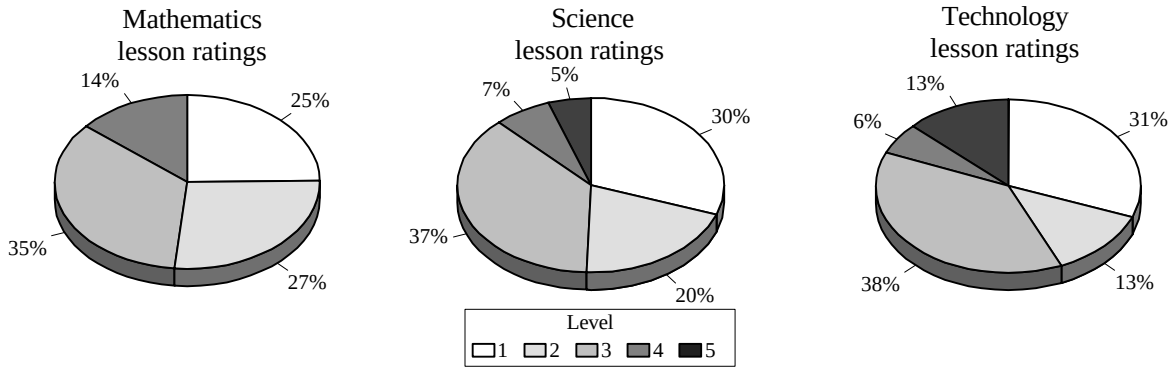


Differences Between Math, Science and Technology Lessons

There were not strong differences between math and science lessons in terms of their overall ratings. However, it is interesting to note that we saw no math lessons that warranted the highest rating of Level 5. One third of the lessons in both disciplines were rated as Level 3.

Researchers rated technology lessons most favorably overall. Eighteen percent of all technology lessons were rated a 4 or 5. It is important to note that all of these "effective" lessons were "technology education" lessons. These technology education classes were often distinguished by their student-centeredness: students were typically involved in design and research, working on their own but facilitated by able teachers.

Overall quality of the lessons by subject

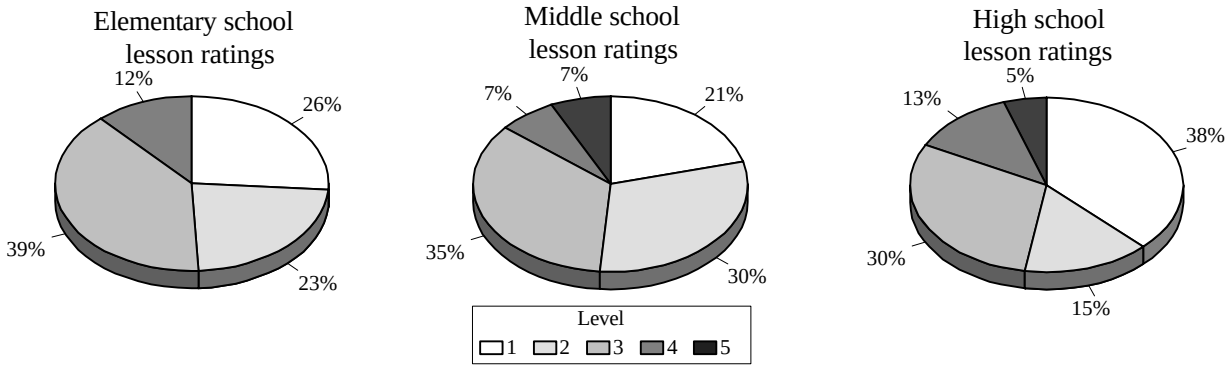


The Overall Quality of Elementary, Middle, and High School Lessons

In examining the ratings for elementary, middle and high school levels more closely, some interesting findings emerged. At the high school level, on one end of the spectrum, we saw many lessons that were rated as passive instruction. These data correspond with our observations that many of the high school Regents classes (advanced level classes which culminate in year-end Regents examinations) had a strong potential to create passive learning environments – these were classrooms characterized by a kind of "forced march" through an arduous curriculum. At the other end of the spectrum, 18% of all high school lessons were seen as accomplished to exemplary; of that 18% some were Regents classes but most were non-Regents classes. Also, many of the high school technology education (non-Regents) classes were highly rated.

Interestingly, we saw no level 5 ratings at the elementary level (though we did see level 4s). The dearth of well-designed student-centered curriculum at the elementary level in the districts we observed likely contributed to this finding. The absence of level 5 ratings may also be an indicator of the fact that elementary teachers often lack the in-depth mathematics and science content knowledge needed to instill deeper meaning into a lesson. Perhaps also indicative of a lack of content knowledge is that many of the elementary classrooms that were rated "1" often involved activity that lacked conceptual purpose and focus.

Overall quality of the elementary, middle, and high school lessons



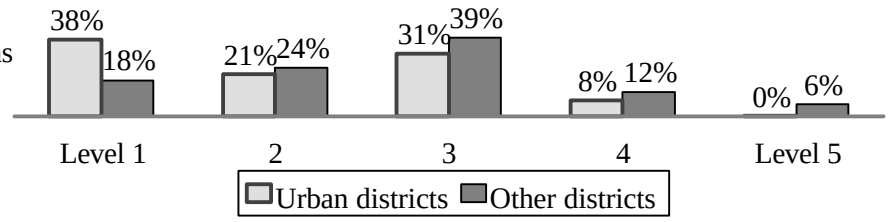
Differences Between Types Of Districts

In addition to examining lesson ratings across the seven districts as a whole, we also looked to see to what extent, if any, the overall quality of lessons varied by the type of districts. We made three comparisons: high-needs communities vs. other communities; urban areas vs. non-urban areas; and small cities vs. larger and smaller communities. Several districts fit in more than one of these categories.

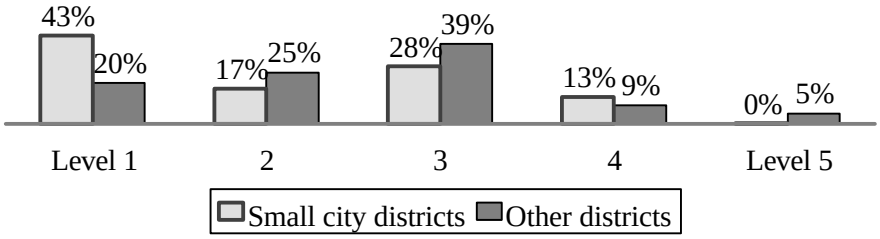
The quality of lessons varied considerably *within* each district we visited. That is, in no district did we find only poor-quality lessons, or only lessons of excellent quality. In general, the variation within a district was greater than the variation across districts. Nonetheless, there were some significant differences between the urban and non-urban districts we studied. Urban classrooms (which include some in New York City, as well as some small cities) had a far higher number of "1" ratings and an absence of exemplary "5" ratings.

Quality of lessons in different types of districts

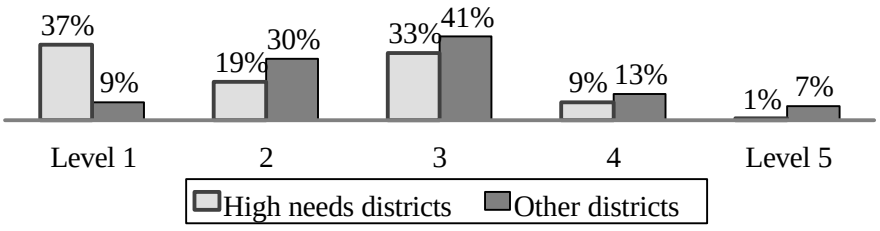
Quality of MST lessons in urban districts compared to MST lessons in other districts



Quality of MST lessons in small city districts compared to MST lessons in other districts



Quality of MST lessons in high-needs districts compared to MST lessons in other districts

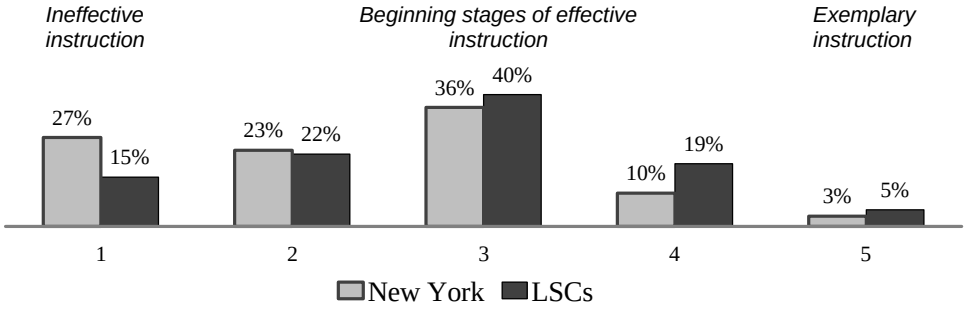


The problems of the large urban and high-needs districts are well known. What was surprising to us was the degree to which we found similar conditions in the state's small cities.

Comparison with HRI National Ratings

We compared this study's classroom ratings with the national pool of classrooms rated as part of the NSF LSC evaluation effort in mathematics and science. It is important to note that the majority of LSC observations are conducted in districts that have NSF funding, and classrooms where the teacher has had 20 or more hours of professional development and is using standards-based curriculum materials. Thus it is perhaps not surprising that overall, the ratings of New York lessons are rated lower than the ratings of lessons in LSC classrooms across the country. While we found that 13% of lessons in New York were rated to be accomplished or exemplary (Level 4 or 5), nearly twice as many of the lessons observed in LSC districts (24%) were rated that highly. The number of lowest ratings (Level 1) of New York lessons is nearly double the national LSC average.

Comparison of the overall ratings of the quality of mathematics and science lessons in New York and in NSF's Local Systemic Change districts

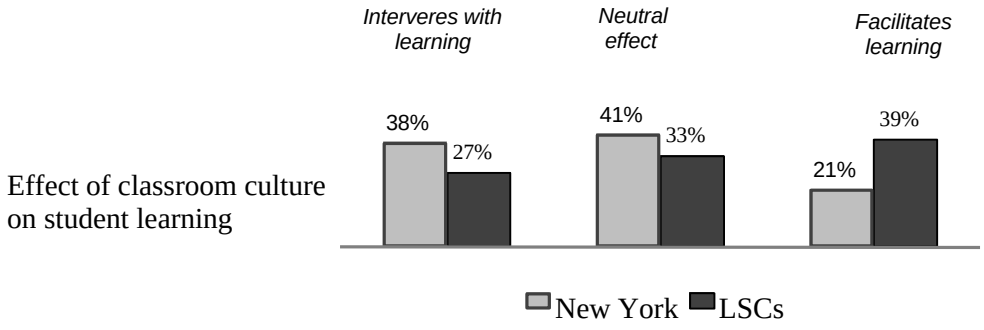


Comparison of Classroom Cultures

The trend continues when looking at classroom culture. Overall, 39% of the LSC lessons were rated a Level 4 or 5 for classroom culture; in New York State 21% of the classrooms' culture was given a Level 4 or 5 rating. These data also point out that the quality of MST instruction depends heavily on the underlying quality and culture of the schools themselves. It should be

noted that the classrooms we observed not only lagged behind the LSC classrooms, but that the deeper culture of teaching and learning in New York classrooms is also quite different from the national sample.

The extent to which classroom culture facilitates learning in New York and LSC classrooms



Summary and Comment

Overall, our ratings of 156 classrooms in New York State show that much of the instruction is at what the Horizon Research protocol terms "the beginning stages of effective implementation." We saw some quite good traditional instruction that clearly is focused on state curricular goals. We found even more instruction that might be called "bland." The reasons for this "blandness" are multiple – some teachers appeared to lack the math and science knowledge to push lessons to a deeper level. Some teachers seemed resigned to follow a very structured curriculum without paying much attention to the thinking and interests of their students. And, in some of the high-needs/small city districts, the conditions were difficult; teachers simply seemed fatigued.

The vision of instruction painted in national standards documents and in the state's Learning Standards for MST is one that is rich with inquiry and personal relevance, as well as rigor. We did not see much inquiry-based instruction in our visits to classrooms with the exception of teachers who were working outside of the curricular norms. And we also did not see much use of, or even awareness of, innovative or exemplary curriculum. Thus, we see the status of New York's classrooms very much as one might expect to find them at the beginning of a reform effort. While the picture painted by our data may seem troubling, it is important to note that we did see some examples of very good teaching, and we did find many teachers who cared deeply about the welfare of their students, and who aspired to enrich their instruction.

There are, we believe, important lessons for New York State and, indeed, the rest of the country in examining these kinds of data. Our purpose here is ***not*** to prescribe "cures" that will improve this situation. (Indeed the situations we found in classrooms often seemed to point to the presence of too many former "cures".) Rather, we hope that data such as these will provoke thinking and discussion about the current status of teaching MST, as well as the aspired vision of instruction.[11] We believe data such as these can supplement and complement student

[11] When we presented this data to State Department of Education staff members, we found a great interest and openness is discussing the data and their implications. Staff said that systematic data on classroom realities was both valuable and rare.

outcome data, and more systemic data such as attendance, grades etc. Indeed, we would argue that New York and other states need to find mechanisms such as this study so that they can examine the quality of teaching in real classrooms in an ongoing fashion. (The school PQR process in New York State is a very good example of this kind of close look at the culture and quality of schooling.) The direct examination, and discussion of the quality (and qualities) of teaching is rare. And yet we find that it is the culture of the school and the quality of the teachers that parents most care about when they choose a school for their children. We think our assessment and accountability systems should reflect these concerns in a way that is more authentic than the simple reporting of student outcomes. That is why we believe that studies such as this one – that involve the direct study of the teaching in actual classrooms – may be one of the best ways to provide both incentive and guidance for improvements in instruction.