Junlei Li, David Klahr, Stephanie Siler

What Lies Beneath the Science Achievement Gap: The Challenges of Aligning Science Instruction With Standards and Tests

Amongst all *instructional* issues facing science education, the one that exerts the most substantial impact on the lasting achievement gap is the "mile-wide, inch-deep" curriculum, which is created by superficial alignments among standards, tests, and instructional materials.

"If science educators had a dime for every time the phrase 'standards-based' or 'aligned with standards' pops up in science textbooks, instructional product brochures, conference programs, or in-service workshop presentations …"

Diminishing "standards" and "alignment" to overused buzzwords or superficial checklists masks the dire need for truly systemic and operational standards-based alignment in science education. In this article, we report the findings of an ongoing collaborative effort between cognitive researchers and urban science teachers to align everyday teaching with standards, tests, and research-based pedagogy. We begin with an analysis of how the width vs. depth dilemma in science teaching manifested itself in yearly test scores and the achievement gap. We review the problematic issues of alignment among standards, instruction, and assessment. We argue that simply matching standards with so-called "standards-based" materials creates a false sense of comfort in a superficially aligned curriculum. We advocate for schools, districts, even states to undertake the difficult but necessary planning process to create a framework of performance objectives to serve as the critical hinge linking standards, instruction, and assessment. Such curriculum planning must set as its first priority the goals of effectively cutting down the girth of yearly science content while efficiently managing the handoff of students between grade levels.

… simply matching standards with so-called "standards-based" materials creates a false sense of comfort in a superficially aligned curriculum.

Research Context and Data Collection

Our research takes place in three urban parochial schools (> 90% eligible for free and reduced lunch programs and > 95% African American). We use one affluent parochial school as a comparison group (< 10% eligible for free and reduced lunch programs and < 10% African American). Science teachers for 6th through 8th grades in the three urban schools collaborate with the research team in both bi-weekly meetings during the school year and summer workshops. In addition, the researchers learn, observe, and co-teach in the urban classrooms. The comparison school is not directly involved in any intervention efforts.

All schools use the same district-wide curriculum guidelines, though the instructional materials vary from school to school. All science teachers are certified in elementary education with approximately half also certified to teach science at elementary or secondary levels. This teacher profile is comparable to nationwide statistics for science teaching in public schools (National Center for Education Statistics, 2002). All schools are annually assessed using the Terra Nova Comprehensive Test of Basic Skills [CTBS] (CTB/McGraw-Hill, 2001), which includes a 40-item multiple choice assessment for science for each grade level. The parochial district evaluates schools based on the annual tests and exerts administrative pressure on principals and teachers to improve performance. There is no "high-stakes" accountability system (e.g., sanction, merit-pay). The achievement data reported here was collected by obtaining students' answer sheets for CTBS tests from both the urban schools and the comparison school. We analyze test performance by test item rather than relying on the gross subject-level data reported by CTB/McGraw-Hill. In addition, we were able to collect, through interviews, surveys, and in-class observations, a detailed record of what each teacher taught in each school. We connected test data and everyday teaching through item analyses that categorized items by topic area and by cognitive demand (Bloom, 1956).

Science Achievement Gap

What follows is a tale of two gaps:

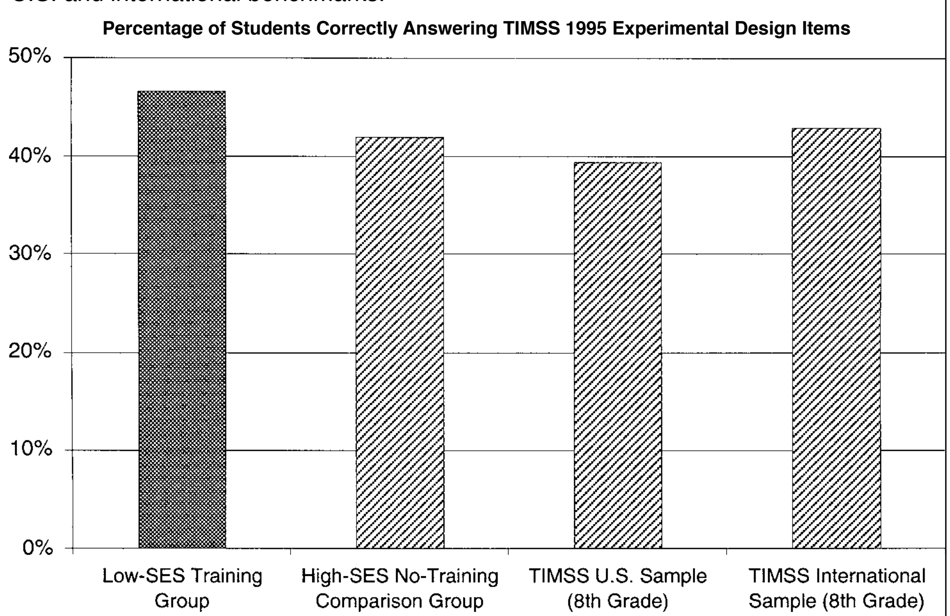
1) The *learning gap* in particular topic areas, which, through teacher and researcher collaboration and intense instructional investment, can be narrowed or even closed;
2) The *test gap* across the entire science curriculum, which, despite teacher and researcher collaboration and intense instructional investment and professional development, remain wide open.

In urban school settings, teaching for mastery requires time and patience.

We began with a set of science instructional strategies first developed in cognitive psychology laboratories and subsequently validated in diverse classroom settings (Chen & Klahr, 1999; Klahr, Chen & Toth, 2001; Klahr & Nigam, 2004; Klahr & Li, 2005; Strand Cary & Klahr, 2005; Toth, Klahr, & Chen, 2000; Triona & Klahr, 2003). For the purpose of our discussion here, the operational details of our instructional method are not particularly important (see Klahr & Li, 2005, for a more detailed discussion). It suffices to say that our proposed methods push for mastery by narrowing our focus on skill or concept domains through a sequence of cognitively-balanced instructional activities, including goal-directed exploration, elicitation of student's justification and explanation, repeated formative and performance assessment, and explicit instruction. The argument we are making here is not that our method is the best way or even that it is better than some other alternative. Instead, we present evidence that our method can close the learning gap while still leaving the test gap wide open.

In urban school settings, teaching for mastery requires time and patience. For example, we had developed instruction to help students achieve high levels of mastery in designing valid scientific experiments. In affluent high-achieving schools, students achieved mastery in two days. In our urban schools, it took one to three weeks depending on classroom and school conditions. But the intense investment of teacher's planning and teaching in urban schools, carried out through iterative lesson studies and in-class teacher-researcher collabora-

Figure 1
Low-SES Training group and high-SES comparison group's performance on select TIMSS 8th Grade science items pertaining to controls and variables, compared with U.S. and international benchmarks.



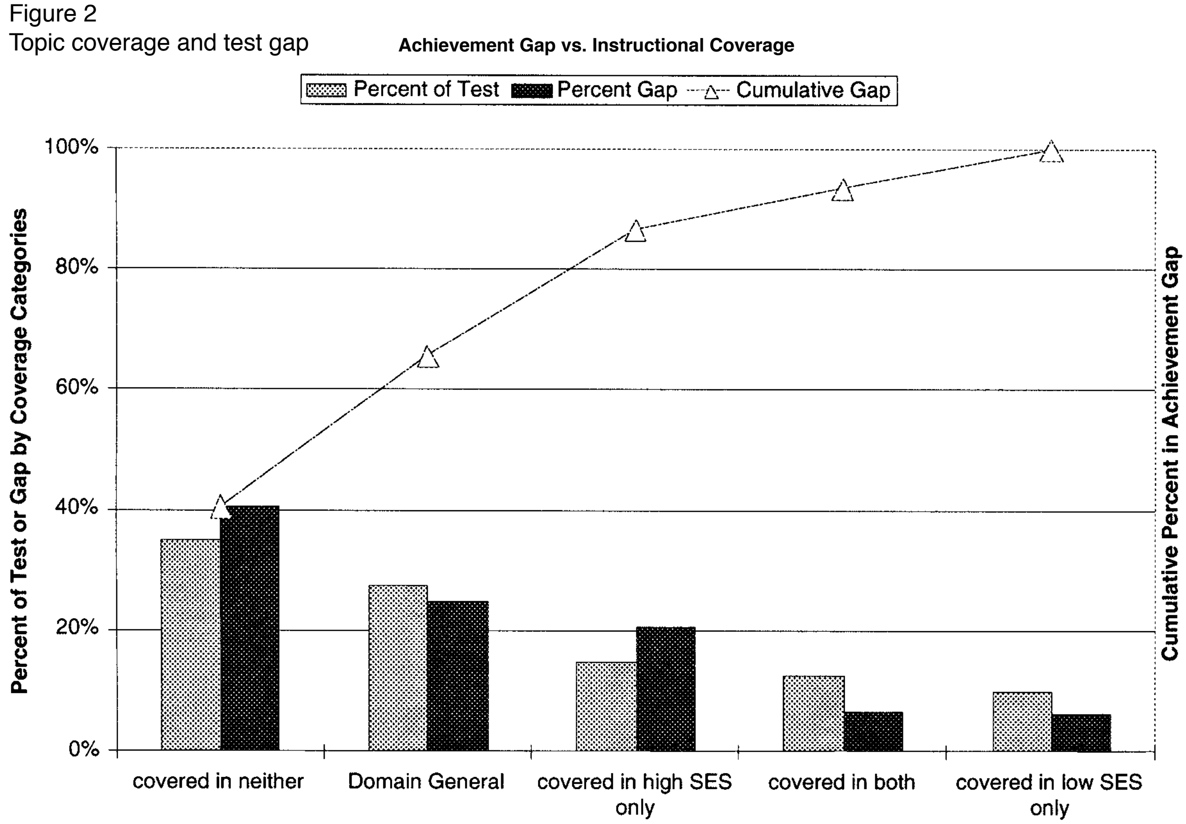
tion, do pay off. In designing scientific experiments, for example, 5th and 6th grade urban students achieved a level of mastery exceeding their same-age counterparts in the affluent school. Their performance also matched or exceeded national and international benchmarks on standardized test items reused from the National Assessment of Educational Progress [NAEP] and Trends in International Math and Science Study [TIMSS] tests (Figure 1). In another example, over a three week period, students in two 6th grade urban classrooms learned to explain day and night and the seasonal change in daylight hours. Their performance on relevant TIMSS 8th grade items not only exceeded that of the U.S. average, but matched that of international leaders like Japan. These results are encouraging indications that, with adequate investment of time, professional development, and research-practice collaboration, we can narrow the learning gap.

One would expect that, with mastery at the topic level, the overall test gap would also narrow. But our efforts did not lead to the narrowing of the gap as measured by yearly standardized tests. The heavy investment in closing the learning gap topic by topic incurs a great cost on the breadth of topic coverage. For each lesson we planned, there were bound to be many that we could not, due to the lack of teacher preparation time. For each topic we taught to mastery, there were bound to be many that we could not, due to the lack of instructional time. In one year, by the time the CTBS test was administered, our three urban schools only managed to cover just over half of the planned curriculum.

The heavy investment in closing the learning gap topic by topic incurs a great cost on the breadth of topic coverage.

The test gap between the urban and comparison schools is illustrated in Figure 2. The 40 items on the test were grouped, based on teacher interviews, surveys, and item analysis, into five categories. The domain-general cat-

Figure 2
Topic coverage and test gap



Note. This parietal chart compares the achievement gap on categories of items based on coverage in low- and high-SES schools. The columns show, respectively, the weight of a particular category of items on the test and the extent to the category contributes to the overall test gap between the schools. The columns are ordered from right to left in terms of the absolute size of the test gap. The lines show the same information but in an accumulative fashion.

egory includes inquiry or reasoning items that do not rely on any specific content knowledge. The remaining categories include items that, without specific content knowledge, a student cannot answer. Figure 2 shows perhaps the "obvious"—when a test item relates to content topics that were taught in the urban schools or skill areas that required no particular content knowledge (i.e., domain general), the associated test gap is smaller per item than that for test items under topics "not taught". This supports our assertion that intense investment in teaching is beginning to narrow the learning gap in the specific topics or skills taught, but not nearly fast enough or "wide" enough to catch up on yearly tests. The test items that fall under topics "not taught" by urban schools contribute to about 60% of the total test gap (adding together the "covered in neither" and "covered in high-SES only" columns). In other words, 60% of the test gap can be attributed not to the quality of teaching in the urban schools, but merely to the breadth of coverage or opportunity to learn. Furthermore, the single biggest source of the test gap is the "covered in neither" category, suggesting that even when both urban schools and the affluent school were limited in their breadth of coverage, differences in prior knowledge alone could account for 40% of the total test gap. It is tempting to jump to the conclusion that breadth of coverage is what we need. But with breadth, we will lose the depth of mastery per topic. During our intervention, the teachers only taught one third of the topics they had taught in past years, but the overall test scores were no different from years prior.

Can we expect this trend to improve over time? Would multiple years of intervention narrow the gap? Though we would like to believe that, based on our success in closing some topic-level learning gaps, our further analysis reveals a more pessimistic answer. Recall that our instruction is focused on mastery and deep understanding. By mastery, we mean that students not only could recognize and reproduce factual information, but could apply their learning robustly in an inquiry context—a goal aligned with the spirit of the standards movements. To what extent is our instructional focus on knowledge application aligned with the assessment instrument? Figure 3 shows the break-down of the 40 test items by cognitive objectives (Bloom, 1956). Over 80% of the achievement gap is contained within the most basic level of Bloom's hierarchy of cognitive objectives, involving mostly terminologies and facts. If we follow the "getting the biggest bang for your buck" principle, we would be tempted to suggest that the quickest path to closing the test gap on the CTBS tests is to target instruction towards the lowest levels of cognitive objectives. This suggests that our instructional focus for understanding and mastery is aligned with the standards but misaligned with the emphasis of the tests.

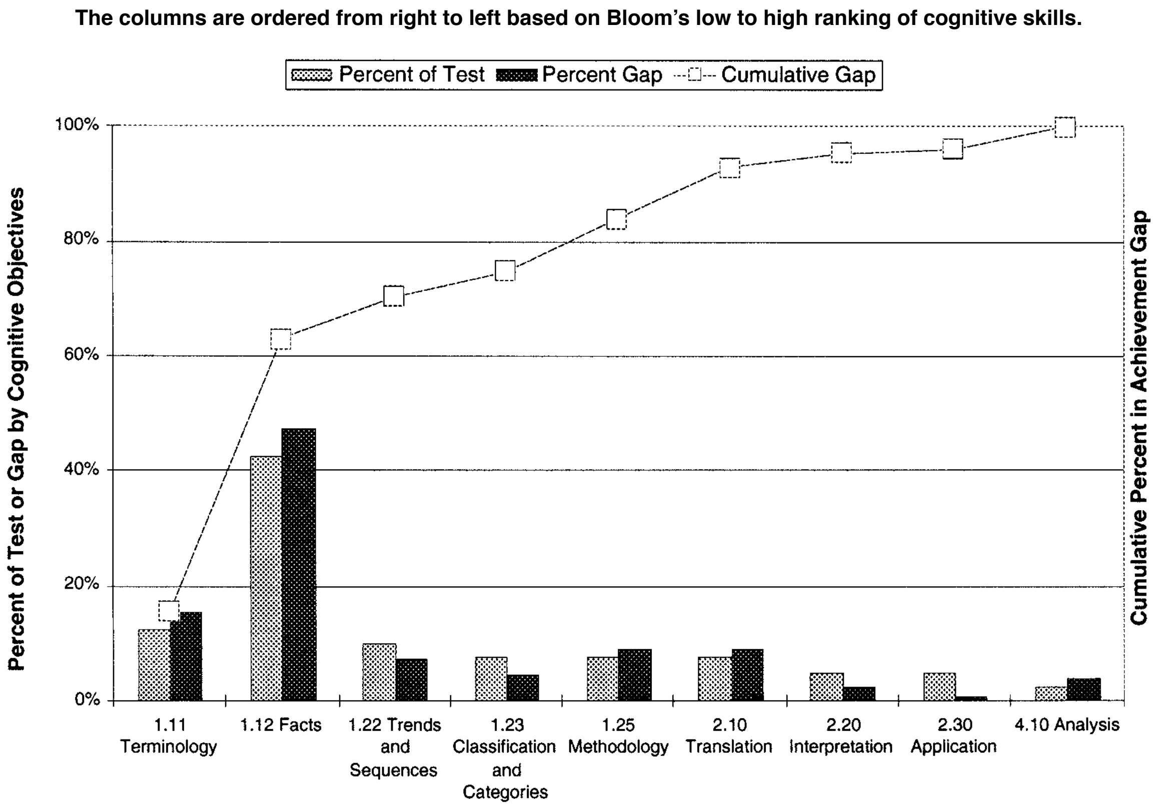
Standards and tests are here to stay and nearly every state has adopted science content standards.

Standards-based Reform in Science Education

We do not infer from the above analysis that schools should do away with standards or tests. Standards and tests are here to stay and nearly every state has adopted science content standards. Beginning in the 2007-2008 school year, all states must also measure science achievement with assessments that align with state standards (No Child Left Behind Act, Public Law 110-107). All of these reform efforts have the intention of narrowing the achievement gap—albeit the gap has mostly been defined as the test performance differences between rich and poor or predominantly white or minority schools.

Before we can operationalize a systemic alignment in science education, we need to first understand the relationship between the achievement gap, standards, tests, and everyday instruction. During the last decade, large scale investigations have focused on international comparisons of test performance, curriculum, and teaching. Two prominent reports—*A Splintered Vision* (Schmidt, McKnight, & Raizen, 1997) and *The Teaching Gap* (Stigler & Hiebert, 1999)—argue that U.S. science and mathematics education are "splintered" by "mile wide, inch deep" curriculum aims and textbooks, and that U.S. teachers have neither the supporting resources nor the ongoing collaborative professional practice to iteratively plan, evaluate, and revise their lessons. The outcry against bloated science curricula and advocacy for professionalizing science teaching are among the core issues that inspired the standards movement (American Association for the Advancement of Science [AAAS], 1990, 1993; National Research Council [NRC], 1999). But is science education less "stuffed" and more "nourished" today than it was more than a decade ago? The debates persist as the standards reform movement, in the course of state-by-state implementation, triggered unintended consequences such as ballooning the

Figure 3
Cognitive objectives and test gap



Note. This parietal chart compares the achievement gap on categories of items based on cognitive objectives (using Bloom's Taxonomy) in low- and high-SES schools. The columns show, respectively, the weight of a particular category of items on the test and the extent to the category contributes to the overall test gap between the schools. The columns are ordered from right to left in terms of the absolute size of the test gap. The lines show the same information but in an accumulative fashion.

scope of science content, limiting the choice of instructional strategies, and imposing one-size-fits-all goals and solutions for diverse student populations (Anderson, 2004; Anderson & Helms, 2001; Barton, 1998; Bauer, 1992; Donmoyer, 1995; Hewson, Kahle, Scantlebury, & Davies, 2001; Li & Klahr, 2005; Rodriguez, 1997; Settlage & Meadows, 2002; Shamos, 1995; Shiland, 1998; Thomas B. Fordham Institute, 2000, 2005; Vesilind & Jones, 1998; Wolk, 1999, 2004).

Academic and policy debates seem somewhat remote to practitioners on the frontline of science education. Science teachers, department heads, and instructional specialists need to survive and thrive in a teaching environment increasingly driven by standards and measured by accountability tests. They are the ones who must, here and now, find solutions to the pressing problems of standards-based reform, including (but not limited to) the following three interrelated claims (Anderson, 2004, p.1):

- The reform agenda is more ambitious than our current resources and infrastructure will support.
- The standards advocate strategies that may not reduce achievement gaps among different groups of students.
- There are too many standards, more than students can learn with understanding in the time we have to teach science.

The analysis of the width vs. depth dilemma in our local context supports each of these three claims. The state of science education, we argue, requires the application of a basic economic principle: *scarcity necessitates choice*. The scarcity of instructional and planning time, physical materials and resources, and teacher preparedness necessitates difficult choices that teachers have to make about what and how to teach on a daily basis. These choices converge on the issue of *alignment*—the streamlining of instructional goals and strategies within policy constraints that maximally utilize the available human, time, and physical resources towards closing achievement gaps.

The outcry against bloated science curricula and advocacy for professionalizing science teaching are among the core issues that inspired the standards movement.

Alignment

There is a broad consensus among practitioners, policy makers, and researchers that "alignment" is a prerequisite for educational improvements in today's high standards and high accountability system (see Olson, 2003 for a succinct summary and advocacy). Everyday instruction in the science classroom should align with standards, be informed by formative and summative assessments that also align with standards, incorporate instructional products that are standards-based, and apply pedagogical strategies that are also standards-based.

What resources are available to classroom teachers, school principals, and district leaders to create such a system of alignment? We are bombarded with documents titled, "content standards", "benchmarks", "teaching standards", and "curriculum frameworks", many of which overlap and restate each other. Educational product brochures are strewn with variants of the "alignment checklist"—generically and superficially claiming how each lesson unit or module is aligned with a host of inquiry and content standards. The overuse and misuse of the term "alignment" belies the genuine alignment process—*to be in or come into precise adjustment or correct relative position* (Webster's Dictionary)—that demands a system in which everyday teaching, standards, and tests can be brought into "correct relative position" through "precise adjustment".

We argue that the lack of an operational process of alignment is not due to the lack of trying, but a dearth of *specificity* and *transparency* in the reform infrastructure. In order to ward off excessive width or depth in teaching, a teacher needs to know *specifically* what content should be taught at what grade level, to what level of mastery, and measured by what set of performance objectives. For example, standards statements like "students should develop general abilities, such as … identifying and controlling variables" (NRC, 1996) and "design controlled experiments, recognize variables, and manipulate variables" (Pennsylvania Department of Education, 2002) could easily have been used to describe goals in undergraduate or graduate level research methods classes. These statements do not offer a usable specification of the level of mastery expected of students in grades 5 through 8. The alternative is to build grade-level performance objective based on the standards, such as:

> "In 5th grade, students should be able to design a controlled experiment when the key variables are already given, in simple topic areas such as, 'Does water make a plant grow faster or slower?' or 'Does sugar dissolve faster in warm water?' In addition, students should be able to discriminate a controlled experiment from an uncontrolled experiment when they are given the variables and the procedures. Also, students should be able to identify the important variable to contrast when the research question has been specified, such as "water" or "temperature of water" in the two topic examples above."

In order to align day-to-day teaching and formative assessment with yearly accountability assessments, a teacher also needs a *transparent* roadmap that leads from topic-specific performance objectives to the skills and knowledge demanded by accountability tests. This roadmap should make it clear and unambiguous what the mandated test expects of the student within a specific topic area at a specific grade level, not some general descriptions of "proficiency". Most states and test publishers release teacher's guides and assessment handbooks in the hopes of providing such a roadmap. But guideline statements often are just as vague and generic as those in the standards, for example:

> "Students must have an understanding of the concepts and terms included in the standards through grade 7. This understanding should go beyond simple knowledge recall (Bloom's Level One). Students should be able to translate and apply the terms to new situations when answering an item." (Pennsylvania Department of Education, 2002)

Using our example earlier, how would a teacher know, from this general assessment guideline, what level of performance is expected from the students when it comes to controlling variables and designing experiments? The alternative is to provide a topic-level roadmap so that the teacher can clearly see the linkage (i.e., transparency) between the standards, the performance objective, and the test requirements. It may look something like this for our example topic:

> "At a 'recall' level, students can define the words 'variable' and 'control'. At a 'basic use' level, students can identify the target variable from a question statement, such as, 'Does water make the plant grows faster?' At an 'application' level, students can design an experimental procedure based on the variables they can identify from a question statement. At a 'generalization' level, students can examine a given description of an experimental procedure and critique whether the procedure has met the requirements of a controlled experiment. For each of these levels, example assessment items are included. At the 4th grade level, assessment items will emphasize recall and basic use. At the 7th grade level, assessment items will emphasize application and generalization."

The state of science education, we argue, requires the application of a basic economic principle: *scarcity necessitates choice.*

These two aspects of alignment, specificity and transparency, cannot be implemented independently. Without specificity of content and performance standards, there is no framework to which the tests or teaching could align. Without transparency in the tests, the outcome measures can only produce information of a coarse grain size, unusable to inform and improve everyday teaching. We believe that, as a prerequisite for improving achievement, we must have a system of alignment that can reduce the burden of the "mile-wide" content and enable meaningful and mastery-focused teaching. This is easier said than done. Using our local context, we review the challenges of using traditional approaches and existing resources to attempt this daunting task.

Difficulty of Alignment In a Local Context

In the same year as our project began, our parochial district unveiled its newly revised curriculum guidelines based on the adoption of the state science content standards. Because the CTBS tests used by the district proclaim to be aligned with science standards at the national level, we evaluated whether the Pennsylvania state standards align with National Science Education Standards [NSES] (NRC, 1996) and the Benchmarks for Scientific Literacy (AAAS, 1993). We assembled all of the standards pertaining to the middle grade levels (5th through 8th). From this collection of content standard statements from three separate guidelines, we group similar topics together into "clusters"—each containing

Table 1
An example of content standard topics used in the alignment analysis

	NSES 5-8	AAAS 6-8	PA 7th
Light & Solar Energy	• Light interacts with matter by transmission (including refraction), absorption, or scattering (including reflection). To see an object, **light from that object--emitted by or scattered from it--must enter the eye.** p155 • The **sun is a major source of energy** for changes on the earth's surface. The sun loses energy by emitting light. A tiny fraction of that light reaches the earth, transferring energy from the sun to the earth. The **sun's energy arrives as light with a range of wavelengths, consisting of visible light, infrared, and ultraviolet radiation**. p155	• Something can be "seen" **when light waves emitted or reflected by it enter the eye**... 4F p90 • Human eyes respond to only a narrow range of **wavelengths of electromagnetic radiation—visible light**. Differences **wavelength** of within that range are perceived as differences in color. 4F p90 • Light from **the sun is made up of a mixture of many different colors of light**, even though to the eye the light looks almost white. Other things that give off or reflect light have a different mix of colors. 4F p90	• Explain how…**light travels** in waves of differing speeds, sizes and frequencies 3.4.7C • Explain how convex and concave mirrors and lenses change light images. 3.4.7C • Know that the **sun is a major source of energy that emits wavelengths of visible light, infrared and ultraviolet radiation**. 3.4.7B

Note: AAAS states that students should "learn about the electromagnetic spectrum, including the assertion that it consists of wavelike radiations. Wavelength should be the property receiving the most attention but only minimal calculation." (p 90)
Copyright: All original written materials are copyrighted by the National Research Council, American Association for the Advancement of Science, and the Pennsylvania Department of Education. We added grouping and re-organization of the original contents.

Table 2
Topics identified across three content standards, NSES, AAAS, and PA

	Earth Science	Life Science	Physical Science
Cluster I	Earth Composition, Plate Tectonics & Related Processes	Structure & Function of Cells	Physical Properties & Phases of Substances
Cluster II	Erosion & Deposition	Levels of Organization & Development	Chemical Changes & Reactions
Cluster III	Rock Cycle & Soil Formation	Human Body Systems	Elements & Compounds
Cluster IV	Natural Resources & Environment	Disease	Motions & Forces
Cluster V	The Atmosphere	Reproduction	Forms & Transfer of Energy
Cluster VI	Water	Heredity	Sound Energy
Cluster VII	Oceans, Climate & Weather	Response, Behavior & Adaptation	Light & Solar Energy
Cluster VIII	Planetary Characteristics & Composition	Populations & Ecosystems	Electricity & Magnetism
Cluster IX	The Universe	Energy use in Ecosystems	
Cluster X	Gravity & Movement in the Solar System	Classification of Organisms	
Cluster XI	Seasons	Extinction & Fossil History	

and comparing relevant statements from all three standards. Table 1 shows one example of such clusters and Table 2 shows the total of 30 clusters identified in the three main branches of middle school science across three sets of standards. We have not included the inquiry standards in our analysis with the understanding that students should be engaged in inquiry across all science content areas.

The three sets of standards, for the most part, ask for a similar core body of content. At least on a content level, we seemed to have found alignment among district, state, and national standards. But in practice, a curriculum plan requires a level of specificity that the standards fall far short of providing. We discuss two significant challenges in curriculum planning using standards: sequence and selective emphasis.

Unlike a curriculum plan, standards provide neither transition between topics nor progression within topics for each block of grades (e.g., 5th through 8th). Using our example topic Light and Solar Energy (Table 1), where does it fit sequentially within the whole spectrum of science content (Table 2)? In addition, which aspects of this topic should be taught in 5th grade vs. 6th grade vs. 7th grade vs. 8th grade? The content standards offer no such specification, leaving this enormously complex task to practitioners. Also unlike a curriculum plan, standards tell what topics should be taught, but not the appropriate emphasis or weight one should place upon different topics at different grade levels.

Without specificity of content and performance standards, there is no framework to which the tests or teaching could align.

Obviously, content standards leave much to do for teachers and science instruction specialists. But on what research, knowledge, and practical grounds should such complex decisions be made?

In the absence of a specific curriculum framework, the teachers in our local context rely heavily on existing instructional materials—including textbooks, lab kits, and miscellaneous activities they have attempted in past years. Much of the materials published after the release of the national standards proclaim their alignment with content and inquiry standards. If materials do indeed align with standards, then why not just follow their predefined sequence and emphases? We could keep our fingers crossed that what one teaches based on standards-based materials matches what one's students would be measured on by the standards-based tests.

The alignment between popular instructional materials and science standards has been extensively studied, particularly in middle school science (Kesidou & Roseman, 2002; Stern & Ahlgren, 2002; Stern & Roseman, 2004). Across the board, popular instructional materials fail to convey the "big ideas" intended by the standards and to provide meaningful assessments appropriate to the knowledge level demanded by the standards. We do not re-investigate these issues, but rather, focus on three commonsense practical issues. First, do the textbooks "cover" the topics in the standards? Using 6th grade as an example, we find that the textbook covers or touches upon 24 of the 30 total clusters (Table 2). The textbook is divided into 59 lesson units, which, if divided by the available school days in a year, require on average 2.5 class periods each. How much content is included in one single lesson unit that is to be taught in 2.5 class periods? Using heredity as an example, the textbook lesson unit contains the following concepts—traits, DNA, gene, Watson & Crick, DNA base types, DNA structure, copies, and ladder, the Human Genome Project, and the use of DNA in police work. Though standards should in theory help us narrow down the coverage, the lack of specificity in the language of the standards invariably favors *inclusion* rather than exclusion of topics. One can quite easily make a case that all of the concepts listed above fall under the relevant state standards, which include statements such as, "know that every organism has a set of genetic instructions", "identify and explain inheritable characteristics", "describe how traits are inherited", "recognize that mutations can alter a gene", and "describe how … genetic technologies can change genetic makeup". Lest these topic-level statements not be inclusive enough, there are always some "catch-all" topic-general standards under broad headings such as, "Science, Technology, and Human Endeavors", with inclusive statements like, "explain how human ingenuity and technology resources satisfy specific human needs and improve the quality of life." The lack of specificity in standards all but ensured that the textbooks will always "cover" standards-based topics.

Second, do the test items align with the topics in the standards? From the CTBS tests, we identified all test items that demanded specific content knowledge and matched them with appropriate topics (inter-rater reliability 85%, disagreements resolved by consensus). All of the content-based items in the 6th grade science test in CTBS fall within 16 of the 30 total clusters. This alignment between test and standards is expected given the general "inclusiveness" of the standards language. For example, the 6th grade test included two test items in the general topic area of "gravity and movements in the solar system". One item asks about the causes of tide and the other compares all nine planets' orbiting times. Easily, the topic and level of these two items align with the standards. The problem is, so would many other possible test items. What about the causes of day and night, summer and winter, sunrise and sunset, changes in length of daylight, or the comparisons of gravitational force on each planet and the moon? How does a teacher know which of these many topics need to be taught deeply when there is no time to teach all of them equally in-depth? None of these ideas are trivial, by any means. The famous "Private Universe" video shows how Harvard graduates and faculties stumble on these supposedly "middle-school" science questions.

As our nation's science education crosses the threshold of accountability testing, it is imperative to build, at whatever level feasible—by state, by district, by school, or by science department if need be—a coherent and operational system of alignment among everyday teaching, content standards, and assessments.

This leads to our last question—does the instructional material used in a particular year cover what is needed to perform on the test items used for the same year? This would seem highly unlikely considering that the textbook is published before the test was ever made and by a different publisher. But like magic, the majority of the test items fall within the topics covered by the textbook (Table 3). Both the textbook and the test seemed to have passed the muster of "standards-based" alignment. Can we simply follow the instructional materials and be confident that, if we teach by these materials, our students would achieve on these aligned tests?

Based on our in-class observations and interviews with 14 science teachers across 6 schools within the parochial district, we hear one unanimous message from all teachers: "I don't have enough time to teach everything. I start slow but then have to rush things through and try to get as much done as I can." Referring back to Table 3, it is easy to see why this would happen. The 30 content topics are meant for all four grade levels from 5th through 8th grade. They are not designed to be taught in a single grade level. The 6th grade textbook, for examples, covers 24 of the 30 clusters. This is the amount of coverage of all general science textbook we surveyed, regardless of grade levels. So teachers are repeating many topics year after year, yet each time could not afford to spend more than a few class periods on each lesson unit. This echoes the depictions of the "mile wide, inch deep" curriculum in the TIMSS report on U.S. Science Education (Schmidt, McKnight, & Raizen, 1997). Nearly a decade after NSES and seven years after "A Splintered Vision", we see ghosts of pre-standards days materialize in our schools, or perhaps, they never left. In this system, we may be able to speak of

Table 3
The alignment among textbook, test, and content standards in 6th Grade

Total of 30 Content Clusters (Grades 5-8, see Table 2)	**Covered in textbook**	**Not covered in textbook**
Tested in CTBS	13	3
Not tested in CTBS	11	3

Table 4
CTBS test coverage in life science across four grade levels

	LIFE SCIENCE	**GRADE 5**	**GRADE 6**	**GRADE 7**	**GRADE 8**
Cluster I	Structure & Function of Cells		1		1
Cluster II	Levels of Organization & Development				
Cluster III	Human Body Systems		2		1
Cluster IV	Disease				1
Cluster V	Reproduction		1	1	1
Cluster VI	Heredity			2	1
Cluster VII	Response, Behavior & Adaptation	2		1	1
Cluster VIII	Populations & Ecosystems	2	3	1	1
Cluster IX	Energy use in Ecosystems	1		1	2
Cluster X	Classification of Organisms	4	2	3	
Cluster XI	Extinction & Fossil History				

Note. Numbers in cell represent the number of test items per year and the corresponding initial weight on the curriculum plan.

alignment and coverage, but certainly not mastery and understanding.

Specificity, Transparency, and Professionalism

In this article, we presented our search for alignment and its problematic relationship to the persisting test gap. We argue that, amongst all *instructional* issues facing science education, the one that exerts the most substantial impact on the lasting achievement gap is the "mile-wide, inch-deep" curriculum. This problem is created by superficial alignments among standards, tests, and instructional materials. It squashes opportunities to innovate, experiment, and plan. As our nation's science education crosses the threshold of accountability testing, it is imperative to build, at whatever level feasible—by state, by district, by school, or by science department if need be—a coherent and operational system of alignment among everyday teaching, content standards, and assessments. Such a solution needs to account for and address the issues of specificity and transparency we have raised. Rhetorical arguments and marketing slogans are simply not useful in the search for such an alignment. We need to do the grunt work. We need to plan lessons topic by topic, measure progress assessment by assessment, and track performance grade by grade, in order to narrow the achievement gap using our scarce resources and ever more precious time.

Acknowledgements

The research described in this article is funded through the Cognition and Student Learning Program at the Institute for Education Sciences, U.S. Department of Education. We thank all the science teachers in six Pittsburgh parochial schools who opened their classrooms for our research. We thank them particularly for their patience in searching for a workable solution amidst the most challenging instructional environment.

Resources

- The complete set of analytical tools and alignment manuals described in this article will be available at http://www.psy.cmu.edu/lessonplans
- The complete set of released TIMSS items and benchmarks from 1995 to 2003 are available at http://timss.bc.edu. The complete set of released NAEP items, scoring sheets, and benchmarks from 1996 and 1999 are available at http://nces.ed.gov/nationsreportcard/itmrls/.

References

American Association for the Advancement of Science (1990). *Science for all Americans*. New York: Oxford University Press.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Anderson, C. W. (2004). Science education research, environmental literacy, and our collective future. *NARST News*, 47 (2). National Association for Research in Science Teaching.

Anderson, R. D., and J. V. Helms. (2001). The ideal of standards and the reality of schools: Needed research. *Journal of Research in Science Teaching*, 38 (1), 3-16.

Bauer, H. (1992). *Scientific literacy and the myth of the scientific method*. Urbana & Chicago: University of Illinois Press.

Barton, A. C. (1998). Reframing "science for all" through the politics of poverty. *Educational Policy*, 12, 525-541.

Black, P., and D. William. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*, 80 (2), 139-148.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives Handbook 1: Cognitive Domain*. New York: Longman, Green & Co.

Chen, Z. and D. Klahr. (1999) All Other Things being Equal: Children's Acquisition of the Control of Variables Strategy. *Child Development*, 70 (5), 1098 - 1120.

CTB/McGraw-Hill (2001). *Terra Nova CAT complete battery plus level 15, form C*. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill (2001). *Terra Nova CAT complete battery plus level 16, form C*. Monterey, CA: CTB/McGraw-Hill.

Donmoyer, R. (1995). The rhetoric and reality of systemic reform: a critique of the proposed National Science Education Standards. *Theory into Practice*, 34 (1), 30-34.

Hewson, P., J. B. Kahle, K. Scantlebury, and D. Davies. (2001). Equitable science education in urban middle schools: Do reform efforts make a difference? *Journal of Research in Science Teaching,* 38, 1130-44.

International Association for the Evaluation of Educational Achievement (1998). *TIMSS science items: Released set for population 2 (seventh and eighth grades)*. Retrieved on September 16, 2004 from http://timss.bc.edu/timss1995i/TIMSSPDF/BSItems.pdf

Kesidou, S., and J. E. Roseman. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, 39(6), 522–549.

Klahr, D. and J. Li. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, 4, 217-238.

Klahr, D. and M. Nigar. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661-667

Klahr, D., Z. Chen, and E. Toth. (2001). Cognitive development and science education: Ships passing in the night or beacons of mutual illumination? In Carver, S. M. and Klahr D. (Eds.) *Cognition and Instruction: 25 years of progress*. Mahwah, NJ: Lawrence Erlbaum Associates.

Li, J. and D. Klahr. (2005). The psychology of scientific thinking: Implications for science teaching and learning. In J. Rhoton & P. Shane (Eds.) *Teaching Science in the 21st Century.* National Science Teachers Association and National Science Education Leadership Association: NSTA Press.

National Center for Education Statistics (2002). *Qualification of the public school teacher workforce: Prevalence of out-of-field teaching 1987-88 to 1999-2000.*

National Center for Education Statistics (n.d.). *The nation's report card (NAEP): 1996 assessment science public release grade 4*. Retrieved on September 16, 2004 from http://nces.ed.gov/nationsreportcard/itmrls/sampleq/96sci4.pdf.

National Center for Education Statistics (n.d.). *The nation's report card (NAEP): 1996 assessment science public release grade 8*. Retrieved on September 16, 2004 from http://nces.ed.gov/nationsreportcard/itmrls/sampleq/96sci8.pdf.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Olson, L. (2003). Standards and tests, keeping them aligned. *Research Points*, 1 (1) 2003. Washington, D.C.: American Educational Research Association.

Pennsylvania Department of Education (2002). *Academic standards for science and technology* and *Academic standards for environment and ecology*.

Rodriguez, A. (1997). The dangerous discourse of invisibility: A critique of the National Research Council's national science education standards. *Journal of Research in Science Teaching*, 34, 19-37.

Schmidt, W. H., C. C. McKnight, and S. A. Raizen. (1997). *A splintered vision: an investigation of U.S. science and mathematics education*. Boston/Dordrecht/London, Kluwer Academic Press.

Settlage, J. and L. Meadows. (2002). Standards-Based Reform and Its Unintended Consequences: Implications for Science Education within America's Urban Schools. *Journal of Research in Science Teaching*, 39, 114-127.

Shamos, M. H. (1995). *The myth of scientific literacy*. New Brunswick, NJ: Rutgers University Press.

Shiland, TW (1998). The atheoretical nature of the national science education standards. *Science Education*, 82, 615-617

Stern, L., and A. Ahlgren. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching*, 39(9), 889–910.

Stern, L., and J. E. Roseman. (2004). Can middle-school science textbooks help students learn important ideas? Findings from Project 2061's curriculum evaluation study: Life science. *Journal of Research in Science Teaching*, 41(6), 538–568.

Stigler, J.W., and J. Hiebert. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Strand Cary, M. and D. Klahr. (2005). Two roads diverged in the classroom, but did it make a difference? Path independence in learning & transfer. (Cognitive Development Society, Biennial Meeting, 2005, San Diego, CA)

Thomas B. Fordham Institute. (2000). *The state of state standards*.

Thomas B. Fordham Institute. (2005). *The state of state science standards*.

Toth, E., D. Klahr, and Z. Chen. (2000). Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition & Instruction*, 18(4), 423-459.

Triona, L. M. and D. Klahr. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition & Instruction*, 21, 149-173.

Vesilind, E.M. and M.G. Jones. (1998). Gardens or graveyards: Science education reform and school culture. *Journal of Research in Science Teaching*, 35 (7), 757-775.

Wolk, R. A. (1999). Making mid-course correction in standards-based reform. *1999 National Education Summit Briefing Book*. Achieve Inc.

Wolk, R. A. (2004). Perspective: Way off course. *Teacher Magazine*, 6 (2), 5

Junlei Li is a postdoctoral fellow at the Department of Psychology at Carnegie Mellon University. He investigates the practical implications of educational policy and cognitive research by co-teaching with teachers in urban school science classrooms. Correspondence concerning this article can be sent to junlei@andrew.cmu.edu.

David Klahr is professor of Psychology at Carnegie Mellon University. He has written many articles and books on the analysis of complex cognitive processes in such diverse areas as voting behavior, college admissions, consumer choice, peer review, problem solving, and scientific reasoning, and has more recently worked on how to better teach children to design and interpret simple experiments.

Stephanie Siler is a postdoctoral research fellow at the Department of Psychology, Carnegie Mellon University. She investigates motivation, instruction, and conceptual development in science learning.

Karen E. Irving

The Impact of Educational Technology on Student Achievement: Assessment *of* and *for* Learning

The author explores current efforts by educators and policy makers to harness the power of educational technology for both assessment *of learning* and assessment *for learning* in K-12 classrooms.

In this era of accountability brought about through international comparisons that pit the U.S. against other nations and testing requirements resulting from legislation such as No Child Left Behind (U.S. Department of Education (USDOE), 2001) educators, administrators and policy makers require more efficient and effective ways to measure and analyze student achievement data. Moreover, individuals at the local and state levels look to teachers and district leaders to use these analyses to benefit future outcomes. The pressure for accountability affects states and districts as well as classroom teachers. This article explores current efforts by educators and policy makers to harness the power of educational technology for both assessment *of learning* and assessment *for learning* in K-12 classrooms. The first section describes efforts to utilize technology for student achievement testing, an assessment *of learning*. Issues addressed include high-stakes and low-stakes testing, issues for classroom implementation, test preparation programs, special education populations, and adaptive testing. The second section outlines the responses of states, districts and schools to the accountability pressures related to data storage, analysis, reporting and data informed decision making. The last section focuses on the role of technology in networked science and mathematics classrooms where immediate feedback devices provide teachers with formative assessment information about student learning to guide instructional strategies. In a connected classroom, the assessment intends to help students learn. The critical difference between assessment *of learning* and assessment *for learning* lies at the heart of current educational technology use in science classrooms.

The critical difference between assessment of learning and assessment for learning lies at the heart of current educational technology use in science classrooms.

Educational Technology for Testing

The increased testing requirements of the "No Child Left Behind" Act (2001) resulted in serious efforts to develop statewide computer-based testing programs to assess student learning. As of the 2004-2005 school year, 16 states have statewide computer-based testing programs in place while 4 additional states are piloting these programs (Fox, 2005). The rapid response, prompt retake possibilities for students close to the cut-off scores and easily captured data provided by computer-based testing programs provide school districts with several advantages. On a practical level, these programs allow schools to gather data to meet the requirements of federal mandates. More importantly they may provide educators necessary information to individualize learning plans for students. Unlike their more traditional predecessors that often require extensive time for scoring and data distribution, computer-based tests provide immediate feedback on student achievement.

However, statewide computer-based testing comes with a cost. While technology can streamline assessment and provide needed information for documentation related to accountability requirements, schools need computers for students to take electronic tests.

Budget deficits threaten to trump these initiatives. In addition to hardware and software costs, school systems must wrestle with issues of test security to limit potential for invalidating scores on high-stakes tests. Students across the state must take the test in a limited time frame and with similar testing conditions to ensure the fairness of the test. Variations in equipment across schools and districts further complicate the issue. Comparisons of test results from paper and pencil tests versus computer-based tests on modern or out-dated equipment introduce important variables into the evaluation process (Olson, 2003). As a natural consequence of these factors, some states have developed online test preparation programs to help improve student scores. As of 2003, twelve states already had computer-based practice exams available to help student prepare for state-mandated tests (Borja, 2003).

Low-stakes diagnostic computer-based tests offer possibilities for improving student performance without many of the issues related to their high-stakes counterparts. Some educators believe that success on low-stakes tests portends improved performance on high-stakes tests. Further, teachers are the recipients of several advantages of classroom computerized testing such as decreased time grading papers and efficient assessment of student learning. Limitations of this approach (many of which also apply to high stakes testing environments) include the need for classroom sets of computers, difficulties with test security (secure sites that cannot be hacked into by enterprising students), and test sites with easy visibility of computer screens for wandering student eyes. However, as test banks become more sophisticated several limitations have

Some educators believe that success on low-stakes tests portends improved performance on high-stakes tests.

been effectively eliminated. Teachers can now prepare multiple versions of assessment instruments for standard multiple choice, true-false or fill in the blank type questions, but questions that require essay style answers or complex multi-step problem solving items continue to present difficulties for computer-based testing and evaluation systems. The efficiencies of time afforded teachers by computer evaluation of tests are lost for these higher level assessment items (Galley, 2003).

Educational technology introduces both issues and advantages for measuring student achievement in special education classrooms. Advances such as assistive technologies have improved students with disabilities' opportunities to fully participate in the classroom (Hitchcock, Meyer, Rose, & Jackson, 2002; Rose, 2001). The use of computer-based testing technologies extends these possibilities by introducing tools such as spell-checkers or reading machines. For example, hearing impaired students in Oregon may select from English, American Sign Language or both in testing situations. Students with visual impairments or reading disabilities may take tests with text readers in Massachusetts, but the cost and limitations of text reading programs limits this possibility for general education students. Text readers still encounter problems with non standard symbols such as Greek letters or equations found in mathematics or science (Goldstein, 2003).

Exploration of new technology-based assessment systems with special education populations provides educators with insights that promise to benefit both regular and special education students. Indiana's use of annual videotapes of oral reading ability as part of an electronic portfolio assessment represents a creative use of educational technology to provide a unique and highly individualized view of student achievement over a multi-year period. The use of these electronic portfolios has the added benefit of greatly reducing the extensive paperwork needed to document student performance, a common grievance of special educators, and provides special educators the opportunity to easily share IDEA-mandated evidence of student progress with parents (Goldstein, 2003).

Another important area of research and test development is adaptive testing. In these situations, student's previous performance determines subsequent questions providing educators with useful diagnostic information. Adaptive testing allows students to self-pace, provides a suitable challenge for each learner, rewards student effort with immediate feedback, improves test security issues due to the differentiated nature of each sequence of items, and offers the advantage of multimedia options in question design such as the inclusion of text, graphs, images, video and animations (Dunkel, 1999). These forms of testing, however, do not meet the demands of NCLB to assess each student against state established grade level standards. South Dakota's adaptive online testing program, for example, serves as a voluntary resource for schools while educators administer a paper and pencil test to satisfy NCLB legislation (Olson, 2003; Trotter, 2003).

Educational Technology and Data Management

A critical advantage of computer-based testing programs is their readily accessible stores of data. Teachers and policy makers need to access, analyze, and interpret student achievement data if these resources are to guide decision making and strategy selection to improve student learning. Expensive and complex data management systems focused on meeting the reporting requirements of the NCLB legislation offer the promise of easily available student information including demographic information for reporting purposes and student achievement information. In a survey of state officials, 15 states responded that NCLB requirements influenced decisions to obtain more sophisticated data management systems (Hoff, 2005). The North Carolina Window of Information on Student Education or NC WISE represents an example of these new data management systems and the challenges of developing them. The NC-WISE implementation program for this $53 million project called for a two-phase approach over a multi-year time span, which was initially launched in 1999 and is scheduled to be fully operational in 2,264 public schools by 2007. When completed this data management system will capture a wide range of essential student data, integrate classroom, instructional and administrative tasks, provide information directly to teachers, and support school wide functions such as scheduling, transcript management, auditing and student services (Hoff, 2005; Reitz & Winter, 2004).

Similarly, educators anticipate a data management tool in Ohio to aid in preparation of individualized educational plans to help high-school students pass state mandated exit exams. Analysis of student's test results will point to areas of student's strengths and weaknesses and suggest strategies to help students improve their test scores. Another Ohio State data-management longitudinal project creates reports on individual students as they progress from grade 3 to grade 8 (Hoff, 2005). Similar projects have been proposed in both Arizona and Hawaii to expand current student data systems to improve the quality of data and teacher access to data (Hoff, 2005).

The demand for more transparent and student centered uses of data about every aspect of the educational process has increased the necessity for databases that talk easily to each other. In the past, data tended to flow upstream, from local schools to districts to states to the federal government. If schools and teachers expect to use these data to affect school wide or classroom decision making, the data must flow equally well in both directions. Disparate computer systems that do not communicate well with each other interfere with this data flow. In addition, the wide audience for educational data impacts choices about the type of data to collect and the method used for storage. Users representing varied constituencies such as the US Department of Education, state and district policy makers and educational agencies, school administrators, classroom teachers, students, parents and community and corporate partners complicate the situation due to their varying intended uses including meeting NCLB reporting requirements, informing parents and students regarding school choice, guiding administrators and teachers on school curriculum, tracking individual student progress, and helping each of these constituents determine how to invest their time and resources (Palaich, Good, & van der Ploeg, 2004).

As data needs continue to grow and become more complex, the standard issues of cost and available infrastructure are coupled with issues of making a wide variety of data available to many different audiences.

The type of data collected influences its potential impact for decision making. Many argue that the current infrequent standardized high-stakes test data collected by states offer insufficient detail to guide teacher classroom strategy choice (Seltzer, Bryk, & Frank, 1994; Seltzer, Choi, & Thum, 2003; Wainer, Hambleton, & Meara, 1999). The yearly grade or school improvement data mandated by NCLB provide volatile data that do not serve as a reliable indicator of school performance (Kane & Staiger, 2002; Linn & Haug, 2002). In addition, older databases often use events such as enrollment dates or testing schedules as organizational schemes which are not useful for classroom level decision making. Schools need access to individual student records to facilitate student-based decisions. Some states, including Michigan, Indiana, Illinois and Wisconsin have either already built statewide data bases or are in the process of doing so. Concerns about student privacy rights as spelled out in the Federal Educational Rights and Privacy Act (FERPA) (U.S. Department of Education, n.d.) have delayed this process in some cases (Palaich et al., 2004).

In summary, growing accountability pressures to collect, store, analyze and report educational data as well as to generate data-based educational policy provide the force behind the growing use of statewide computer-based testing programs and sophisticated data management systems.

As data needs continue to grow and become more complex, the standard issues of cost and available infrastructure are coupled with issues of making a wide variety of data available to many different audiences. Elected officials, policy makers and educators need sophisticated technical expertise to grapple with the challenges of compatibility and information flow in designing and implementing these data management systems. Evolving educational data needs require thoughtful solutions.

Connected classroom systems offer opportunities for improved formative assessment through questioning and immediate feedback, which has the potential for affording teachers the necessary information to tailor instruction to meet student needs.

Formative Assessment in Connected Classrooms

While the present focus on standardized testing and data collection and management offers both assistance and obstacles related to supporting student learning, standardized testing falls short of providing teachers with appropriate data to support daily decisions in the classroom. Teachers have long struggled to develop mechanisms for monitoring student progress toward understanding and are often frustrated by the quality of information gained through normal classroom processes.

Some have questioned whether the current interest and concern in school accountability can lead to improved student achievement. These issues may also overshadow more complex matters such as educators' efforts to motivate all students to want to learn or to increase students perceptions of their capability to learn (Stiggins, 2002). While the assessment community focuses on ever more sophisticated measures of student achievement and works diligently to ensure reliable and valid measures of student achievement, perhaps another worthy goal would be to maximize the positive impact of test scores on student learning. That is, perhaps these efforts should be dedicated to assessment *for* learning.

A different kind of educational technology use intended to support student achievement occurs in the connected classroom. Connected classroom technology refers to a networked system of personal computers or handheld devices specifically designed to be used in a classroom for interactive teaching and learning. These networked technologies include response systems, classroom communication systems, and newer systems included under the name CATAALYST (Classroom Aggregation Technology for Activating and Assessing Your Students' Thinking) (Roschelle, Penuel, & Abrahamson, 2004). Connected classroom systems offer opportunities for improved formative assessment through questioning and immediate feedback, which has the potential for affording teachers the necessary information to tailor instruction to meet student needs (Black & Wiliam, 1998; Fuchs & Fuchs, 1986). One mode of instruction in connected classrooms involves students beaming answers to a receiving station with an accompanying anonymous display of histograms of student answer choices. Data logs of student responses can be archived for later teacher analysis. Discourse that occurs in a safe environment through the public examination of problem solving and alternative conceptions helps students understand their role as critical listeners and thinkers in the classroom (Artzt & Yaloz-Femia, 1999). In the connected classroom, teacher adaptive expertise forms the critical foundation for formative assessment practices that provide needed information to monitor student's incremental progress and keep them oriented on the path to deep conceptual understanding.

The effectiveness of the connected classroom technology, as with all instructional tools, depends on the skill of the instructor.

Formative assessment coupled with well orchestrated classroom discourse informs teacher practice by providing a window into both what and how students are thinking. Unlike summative assessment that occurs after completion of instruction, formative assessment is ongoing and low stakes. The focus of this kind of assessment is often shifted from the right answer to the reasoning that led to the answer. In a connected classroom, students are informed about their classmates thinking processes, and students with a minority opinion see that they are not alone. Connected classroom environments support positive student thinking habits such as seeking alternative representations for problems, comparing and contrasting different solution strategies, and explaining and describing problem solving strategies. These habits support active engagement and intellectual growth (Dufresne, Gerace, Mestre, & Leonard, 2000).

In a traditional classroom, teachers collect student work, evaluate it by hand, and return the papers to students several days later (Figure 1). In many cases, students have forgotten the point of the assignment and teachers experience difficulty correcting conceptual problems days after instruction on the topic. With the pace of instruction continuing unabated, student naïve understandings accumulate resulting in increasing challenges for student achievement. In science and mathematics where conceptions build incrementally, delayed feedback proves particularly damaging. In connected classrooms, display of aggregated student results occurs shortly after students submit their individual work. Displays can give powerful clues to what students are doing, thinking, and understanding (Roschelle, Penuel, & Abrahamson, 2004). The teacher has immediate information and can use this to adjust instruction. Assignments can be smaller but more frequent—for example, requests to answer a question, solve a problem, state a position, write an equation, give a reason, participate in a simulation, or play a learning game (Stroup et al., 2002; Wilensky & Stroup, 2000, 2002). If the teacher chooses, aggregate class results can be displayed, students receive immediate private and non-threatening feedback in a way that encourages reflections and classroom discourse (Abrahamson, Davidian, & Lippai, 2002; Dufresne, Gerace, Leonard, Mestre, & Wenk, 1996).

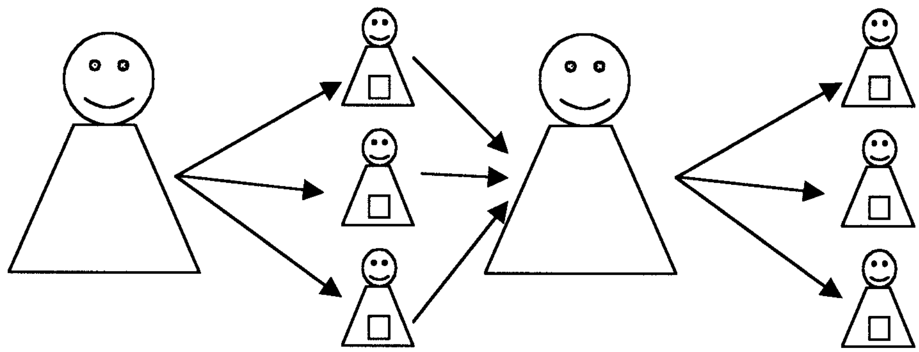
Researchers have found that in a connected classroom, students are more likely to actively engage in the learning process. Student attitudes and motivation improve as teachers focus on conceptual understanding. Quick attention to alternative understanding changes the student experience in science and mathematics from memorization of large quantities of facts to understanding the relationships between concepts (Dufresne et al., 1996).

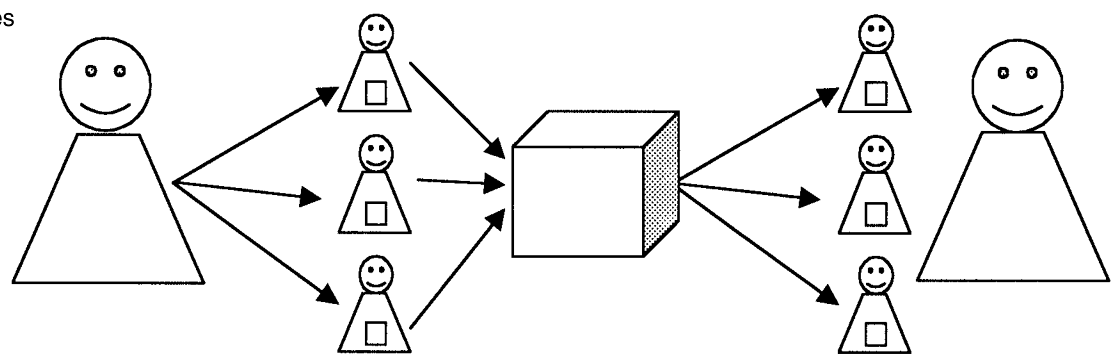
In connected classrooms, display of aggregated student results occurs shortly after students submit their individual work.

Present educational policy focuses on inputs to the educational system—students, teachers, resources, management, standards, and preparation for high stakes test. The assumption behind the input argument rests with the logic that changing these inputs will change the desired outputs—increased student achievement, strategic learners, highly qualified workers, and contented teachers. The pedagogy of the connected classroom addresses a significant component of what teachers and students do inside the black box of their classroom (Black & Wiliam, 1998). Inquiry and problem-solving approaches often leave unanswered the question of what students have learned from these experiences. Teachers need tools to better understand their students' learning *as they engage* their students in learning experiences. While classroom discourse processes have been examined as catalysts for fostering student active cognitive engagement and vehicles for assessing student understanding, the connected classroom offers tools for data collection and display that support these efforts. These tools provide a window into a learner's prior and present understandings and feedback loops that support teacher's instructional decision making and monitoring. Both teachers and students learn during instruction. Through constant observation and monitoring for understanding, teachers and learners may alter the course of the learning experience at critical junctures in the learning process and impact outcomes.

Two important research studies with connected classroom technology are currently underway in the United States: 1) the 5-year NSF-funded (TPC-0456124, 2005) *Teacher Learning of Technology-Enhanced Formative Assessment* project by Leonard, Gerace, Beatty and Feldman at the Scientific Reasoning Research Institute, University of Massachusetts Amherst, and 2) the 4-year, IES funded *Classroom Connectivity in Promoting Student Achievement in Mathematics and Science Achievement* project by Owens, Abrahamson, Pape, Irving, and Demana at The Ohio State University (R305K050045). The Massachusetts team hopes to better understand teacher learning of technology-enhanced formative assessment, to understand effective and efficient methods of teacher professional development for connected classrooms, and to develop tools and techniques for the evaluation of teacher pedagogy in connected classroom environments. The project at Ohio State is an interdisciplinary research study focused on teaching and learning of mathematics and science at the 7th to 10th grade levels with a national sample

Figure 1.
Traditional versus connected classroom





of teachers using the TI-Navigator connected classroom hardware. This project includes a randomized cross-over trial design where the control group is sequentially exposed to the intervention. The project will measure student achievement in algebra 1 and physical science, student motivation, student dispositions toward mathematics and science and self-regulated learning over a four year period for up to 145 teachers and their students. Both of these projects promise to offer scientifically-based evidence to further support the claims made for increased student motivation, engagement, and achievement in connected classroom environments.

Conclusions

Educators, policy makers, parents, and community members wish to maximize student achievement in the United States in science, mathematics and many other disciplines. To achieve this goal, both assessment *of learning* and assessment *for learning* must play important roles. Computer-based statewide testing programs, reliable and valid measures of student learning, and comprehensive data management systems that facilitate the flow of information between stakeholders support the goal of assessment *of learning*. Assessment programs sensitive to the needs of students in both the regular and special education populations provide insight into student learning over time and begin our efforts toward assessment *for learning*. Innovations such as connected classroom technologies, however, offer greater potential for improved formative assessment for classroom teachers on the front lines of assessment *for learning*. In both cases the issue resolves to using reliable data to make decisions, whether at the state, district, or school level, or the classroom teacher level. Educational technology provides a powerful tool to gather, organize, and analyze data. Teachers and policy makers need the expertise to responsibly use these data to benefit all our students.

References

Abrahamson, L., Davidian, A., and A. Lippai. (2002, November 16-19). *Wireless calculator networks -Why they work, where they came from, and where they're going*. Paper presented at the 13th Annual International Conference on Technology in Collegiate Mathematics, Atlanta, Georgia.

Artzt, A., and S. Yaloz-Femia. (1999). Mathematical reasoning during small-group problem solving. In L. Stiff & F. Curio (Eds.), *Developing Mathematical Reasoning in Grades K-12: 1999 Yearbook* (pp. 115—126). Reston, VA: National Council of Teachers of Mathematics.

Black, P., and D. Wiliam. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London.

Borja, R. R. (2003, May 8). Preparing for the big test. *Education Week's Technology Counts 2003, 22,* 23-26.

Dufresne, R. J., W.J. Gerace, W.J. Leonard, J.P. Mestre, and L. Wenk. (1996). *Classtalk: A classroom communication system for active learning* (No. UMPERG Technical Report PERG-2000#09-SEP#1-28pp.). Amherst, MA: University of Massachusetts Physics Education Research Group.

Dufresne, R. J., W.J. Gerace, J.P. Mestre, and W.J. Leonard. (2000). *ASK.IT / A2L: Assessing student knowledge with instructional technology* (No. PERG-2000#09-Sep#1-28pp). Amherst, MA: University of Massachusetts Physics Education Research Group.

Dunkel, P. A. (1999). Considerations in developing and using computer-adaptive tests to assess second language proficiency. *Language Learning & Technology, 2*(2), 77-93.

Fox, E. (2005, May 5). Tracking U.S. trends. *Education Week, 24,* 40-42.

Fuchs, L. S., and D. Fuchs. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*(3), 199-208.

Galley, M. (2003, May 8). The teacher's new test. *Education Week, 22,* 31-33.

Goldstein, L. F. (2003, May 8). Spec. Ed. tech sparks ideas: Testing tools for children with disabilities attracts mainstream attention. *Education Week Technology Counts 2003, 22,* 27-30.

Hitchcock, C., A. Meyer, D. Rose, and R. Jackson. (2002). Providing new access to the general curriculum: Universal design for learning. *TEACHING Exceptional Children, 35*(2), 8-17.

Hoff, D. J. (2005, May 5). NCLB Focuses on data tools. *Education Week, 24,* 12-17.

Kane, T. J., and D.O. Staiger. (2002). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings papers on education policy: 2002*. Washington, D.C.: Brookings Institution Press.

Linn, R. L., and C. Haug. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 29-36.

Olson, L. (2003, May 8). Legal twists digital turns. *Education Week's Technology Counts 2003, 22,* 11-16.

Palaich, R. M., D.G. Good, and A. van der Ploeg. (2004). State data systems that increase learning and improve accountability. *Policy Issues: Research Based Analysis of Policy Based Issues, June*(16), 1-11.

Reitz, R., and C. Winter. (2004). Growing pains: Piloting a statewide information is difficult but worth the effort. *American School Board Journal, September*, 50-51.

Roschelle, J., W. R. Penuel, and L. Abrahamson. (2004). The networked classroom. *Educational Leadership, 61*(5), 50-54.

Rose, D. (2001). Universal design for learning. *Journal of Special Education Technology, 16*(4), 64-67.

Seltzer, M. H., A. S. Bryk, and K. A. Frank. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation & Policy Analysis, 16* (Spring), 41-49.

Seltzer, M. H., K. Choi, and Y.M. Thum. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation & Policy Analysis, 25*(3), 263-286.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83*(10), 758-765.

Stroup, W. M., J. Kaput, N. Ares, U. Wilensky, S. Hegedus, and J. Roschelle. (2002). *The nature and future of classroom connectivity: The dialectics of mathematics in the social space*. Paper presented at the Psychology and Mathematics Education -North America Conference, Athens, GA.

Trotter, A. (2003, May 8). A question of direction: Adaptive testing puts federal officials and experts at odds. *Education Week's Technology Counts 2003, 22,* 17-21.

U. S. Department of Education. (n.d.). *General Family Educational Rights and Privacy Act*, from http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html

U.S. Department of Education. (2001). *No Child Left Behind Act of 2001*. Retrieved October 25, 2005, from http://www.ed.gov/policy/elsec/leg/esea02/index.html

Wainer, H., R. K. Hambleton, and K. Meara. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement, 36*(4), 301-335.

Wilensky, U., and W.M. Stroup. (2000). *Networked gridlock: Students enacting complex dynamic phenomena with the HubNet architecture*. Paper presented at the Fourth Annual International Conference of the Learning Sciences, Ann Arbor, MI.

Wilensky, U., and W.M. Stroup. (2002, April). *Participatory simulations: Envisioning the networked classroom as a way to support systems learning for all*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Karen E. Irving is assistant professor of Mathematics, Science, and Technology in the School of Teaching and Learning at The Ohio State University and co-director of the West-Central Excel Center for Excellence in Science and Mathematics. Correspondence concerning this article may be sent to Irving.8@osu.edu

Mistilina Sato, Vicki Baker, Elaine Fong,
Joni Gilbertson, Tracey Liebig, Neil Schwartzfarb

Changing Mindsets About Classroom Assessment

A group of middle school science teachers and a university researcher recount some of their experiences as they individually and collectively worked toward improving their everyday assessment practices to better support student learning.

Everyday classroom assessment has the unleashed potential to help students improve their performance and deepen their learning. With much public and political attention focusing on summative assessments of what students know and are able to do at benchmark years through state testing programs, relatively little attention has been given to the formative assessment that takes place in the day-to-day interactions between teacher and students and among students. In their review of hundreds of empirical studies of classroom-based assessment, Black and Wiliam (1998) concluded that formative assessment practices are an important and significant contributor to improving student achievement and learning. But how can teachers begin to grapple with their own everyday assessment practices and bring greater attention to using assessment to support learning? This paper describes how a group of middle school science teachers working with a team of researchers from Stanford University in the Classroom Assessment Project to Improve Teaching and Learning (CAPITAL) began to rethink the purposes of assessment and how to use assessment information in their classrooms. The teacher co-authors recount some of their experiences as they individually and collectively worked through a process of change toward improving their everyday assessment practices to better support student learning. Additional findings from CAPITAL have been summarized elsewhere (Atkin et al., 2005; Coffey et al., 2005; Sato et al., 2005). The following experience described by Tracey summarizes a shift in her thinking in the purposes of assessment in her practice.

A Changed View of the Purposes of Classroom Assessment: Tracey's Experience

I just returned from the California Science Teacher's Association (CSTA) Convention. After reflecting on my experience at the convention, I realized I viewed the event through my changed assessment eyes. The summative aspects of assessment have always been clear to me—grading students' work, giving tests at the end of a unit, assigning culminating projects—as a way to know if my students "got it" after all the teaching and instruction were done. I have come to see assessment in my teaching practices as serving a new kind of purpose. The formative side of assessment keeps me focused not only on what students learned after I have taught them, but also on how I can support students' learning while I am teaching them.

While on the train to the convention, two teachers recognized my conference materials and paper grading and joined me for a chat. We exchanged pleasantries as we all graded papers. I was grading an end of unit test. This was the first "traditionally" graded feedback my students would get this school year and it was already the end of October. This assignment was preceded by several topic-specific "questions of the day" that I had stamped when completed and students had corrected; quizzes that had been self-assessed by the students using class examples that showed varied degrees of understanding of the concepts; and a project that had been self, peer, and teacher assessed using a checklist of criteria, then revised and reworked until the criteria for the project had been met by the students. The tests I was scoring showed that the majority of students had met the criteria and they demonstrated understanding of the science standards on which we had focused.

My fellow train riders were not so

pleased with their students' efforts. One mumbled about his students' lack of effort, the other about a colleague who allowed open book tests, and both lamented the lack of parent support for their students' success. Both of them were grading the student work without giving comments to the students. One of the teachers marked the incorrect labels on a skeleton diagram. As I watched this practice, I asked myself, "Will that red mark correct the misnamed bone for the student? What would the marks on the paper and the grade on the top do for the student? Would the student study his mistakes?"

As I asked myself these questions, I realized that I had once marked papers similarly to how I was observing these teachers on the train mark their students' papers. I assigned work, collected it, marked it, and handed it back. What had all of this work on my part, as the teacher, really done for the students? The questions I now asked myself were much less about why the students didn't "get it" after I had laboriously taught them and checked their work. I wondered about how the assignments given to students could be supported with feedback and suggestions for revision that would help students make progress in their understanding and skill development.

As I attended workshops throughout the conference my questions about learning goals and methods of supporting students toward those goals through formative assessment were unleashed. I attended a session about building classroom habitats, since caring for classroom animals and plants appealed to me. The ideas presented in the workshop centered on managing students and materials in cooperative groups. I was struck by how little attention was given to what the students

I wondered about how the assignments given to students could be supported with feedback and suggestions for revision that would help students make progress in their understanding and skill development.

might learn from the activities involving the living animals and plants. The absence of content or learning goals in the workshop conversation felt like a huge gaping hole to me. In my conversations with colleagues, we have come to begin our conversations with what it is we want the students to learn through the activities and experiences we plan in our classrooms.

The presenter brought up peer assessment briefly and this caught my attention. He discussed peer assessment as a process of one student grading another student using an A, B, or C grade; a grade lower than a C should not be given in order to protect students' self esteem according to this presenter. Inside me, the questions bubbled up: "What criteria do the students use to assess one another? What does meeting the criteria look like? Can the students revise their work after the peer assessment? Why do students have to give each other grades? Will the students have an opportunity to discuss their work with one another during the peer review process? How is this peer assessment different from the teacher assigning a grade to the work? Will the students really be fooled by the grading scale that ends at C?" The assessment focus of my questions brought up not only issues of how I would know that my students "got it," but also helped me think about how my instruction could be designed to support the learning while I was teaching.

CAPITAL: An Action Approach to Changing Classroom Practices

The group of teachers co-authoring this paper are from the New Haven Unified School District in Union City, California. They are one group from among the twenty-five teachers who participated in CAPITAL during its four years of National Science Foundation support (NSF Grant REC-9909370). CAPITAL staff and this group of teachers worked together during the 2000-2002 school years, meeting regularly to discuss their current and changing assessment practices through a process of collaborative action research. Teachers used CAPITAL meetings to share classroom practices they were trying and to delve into the underlying reasons guiding their choices. The university staff introduced research findings and ideas from other teachers into the conversations, raised questions for the group or an individual to consider, and participated in the discussions.

Several aspects of classroom-based assessment influenced the work of CAPITAL. Substantial research suggests that greater student learning and higher task performance are achieved by providing task-oriented feedback to students (Butler, 1987; Crooks, 1988), eliciting information from students through assignments and discussion as a means of gauging where students are in their progress toward a goal (Duschl & Gitomer, 1997), and providing opportunities for students to peer- or self-assess their work prior to submitting it for teacher evaluation

(Schunk, 1996). Research on learning also suggests that understanding is strengthened when the learner is asked to take an active part in determining what he or she understands and how he or she came to that understanding through reflection and metacognitive opportunities through talk and writing (National Research Council, 2000; Palincsar & Brown, 1984; Scardamalia, Bereiter, & Steinbach, 1984; White & Frederiksen, 1998).

Substantial research suggests that greater student learning and higher task performance are achieved by providing task-oriented feedback to students

CAPITAL viewed teachers as practitioners who employ practical reasoning, or reasoning directed toward taking principled action, in their professional work (Gauthier, 1963; Schon, 1983). The university staff did not approach the work with teachers as a process of reporting research findings and expecting the teachers to employ specific strategies or techniques in their classrooms. As a group of researchers and teachers working together, we explored how the assessment practices instituted in the classrooms made sense for the differing needs of the students and aligned with the priorities of the teacher-as-person and the context in which the teachers worked. Within the group from New Haven Unified School District, all five teachers expressed fundamental shifts in both their practice and their mindsets about the role that assessment played in their classrooms. The three experiences described below provide illustrative examples of some of the changes in assessment practice and changes in beliefs about assessment. The reader should also note how the teachers describe the dynamic interaction between their actions and beliefs.

Knowing What My Students Are Thinking: Vicki's Experience

Almost every day, I want to know how my students are doing and what they are thinking. One method I use to gather this information from my students is the "question of the day." The question of the day is a question that I put on the board every day and it is meant to either get my students thinking about the concept I am about to teach or to review a concept we have learned the day before.

When it was first introduced to me as a new teacher as a behavior control technique, the purpose was to get students on task as soon as they walked in the classroom door and to provide me with some quiet time to take attendance; the substance of the question or what I would do with the student responses was not discussed. I found the technique useful to some degree at first, but I eventually stopped using it because the questions took too much time out of our very short class periods.

Many years later, I have now started using question of the day again, and I use it for quite a different purpose than classroom management. I choose the questions very carefully so that they are integral to what we are learning in class. I take the time to walk around the room, giving feedback to every student in the class on their answers. Feedback might include comments such as:

> "Good job. Can you help Sean? He seems to be struggling a bit."
> "You calculated the number right, but what units are you using there?"
> "I think I know what you're trying to say, but I am not getting that from your sentence. Try rephrasing it."
> "That's a really creative answer. Would you write that on the board to share with the class?"

I find that the more I use this technique, the better I get at writing good questions and at picking apart the concepts that trouble the students. As an example, in a recent unit on forces for eighth grade, we did a very simple lab using a spring scale to pull a single book across a table, followed by pulling two stacked books across the same table, then three books, and so on. We graphed the data and found that we got a straight line of best fit. The next day, I drew the same graph on the board for the question of the day and wrote: "What does this graph tell you?" I was unpleasantly surprised by the responses.

> "That it is a graph of force and books."
> "That force moved the books"
> "That you need force to move books."

Very few students articulated that the graph meant that the greater the number of books, the greater the force needed to move them. And no one explained that the graph, with its linear relationship, meant that we can determine the force needed to pull any number of books. While I did not expect to see the second idea articulated, I did expect to see the first. I think I would have missed this

The most valuable thing we can do to help with our students' learning is to really be there in class, giving as much feedback as we can directly to the students to support them as they engage in the learning of new concepts and skills.

aspect of the students' understanding if I had not done a question of the day. But now that I knew what the students were thinking, I made a mental note to give at least one more linear graphing opportunity before the unit test. I said to many students, "Well, yes, it is a graph of books and force, but what happens when you add more books?" The problem that I had not anticipated was the use of the word force as used in *Star Wars*: "The force moved the books." That informed my next question of the day, which was, "What's wrong with this sentence?: 'Force moved that chair.'" Before my work with CAPITAL, the students would have completed the lab and I would have expected them to tell me what a graph of force versus books meant on the unit test. With the immediate assessment opportunity that the question of the day offered, I created learning opportunities for my students to interpret a graph thoroughly and to develop a better understanding of the idea of force.

Informing My Instruction: Joni's Experience

I think that I am now giving clearer expectations about learning and creating an equal opportunity for all students to be successful. For me, I have begun to see and use assessment as an integral part of my instruction. I used to think of assessment as something that followed instruction:

instruction → assessment (grading)

I now see it more like a cycle in which assessment outcomes drive my instruction, followed by more assessment:

instruction → assessment → teacher reflection on student learning → instruction → assessment

A recent example from my classroom involved the design of an inquiry-based lab on light using light boxes, prisms, and lenses. Over three days, students investigated the various properties of light. While the students worked with the materials at their lab tables, I was able to walk around and talk with them about their conceptions of light. I frequently encountered misconceptions that we addressed on the spot, I answered the students' questions, and I gave verbal feedback on what I thought they might try next to further their understanding. With the information I gleaned from my one-on-one interactions with the students and lab groups, I used the whole class discussion period to address and clarify some information about the behavior of light and how some of the equipment worked. As a class, we then formulated further questions for investigation.

I am using more analysis of student work to influence my instruction. I see assessment as a tool for me to adjust my instruction and to help students set directions for their learning, and not just as a measure of learning after instruction is completed. From this new perspective on assessment, I now also see that if the instruction is not meaningful and rich, the assessment will not be as reflective of the students' needs. For example, I do not think I would have learned as much about my students' understanding of light if the work they were doing was based on completing a worksheet from the textbook materials. Because of this, I find that I really question my choice of activities and assignments for the students. I ask myself, "How is this assignment meeting the content standards? What do I want my students to learn? What are the key elements that will demonstrate understanding?" I want the opportunity to see and hear them thinking so I really am selective about what I choose to do with them during class time now. I am clear in my mind about the purpose and expected outcomes of the assignment before I ever write a lesson plan.

Evaluating Student Projects: Elaine's Experience

When I assign projects now, I hand out two sheets. The first one gives a clear explanation or instructions for the project. The second sheet is a self-assessment check off sheet. This sheet has everything I am aiming for and looking for in the project. If the student has satisfied all these requirements on his or her paper, he or she will have earned an "A". The self-assessment tool gives the student the opportunity to self-assess the project before giving it to me. The student then knows where he or she stands with regard to the criteria and can revise whatever needs to be revised. In terms of the final evaluation of a student's work, if a student turns in a project that does not yet meet the criteria set out for the project, that student still has the opportunity to make the corrections, turn it in, and still earn full credit, or an "A".

I explained this assessment and grading system of projects being acceptable (i.e., they meet all the criteria) or unacceptable (i.e., the project needs revision) to parents at Back to School Night and they loved it. So far I have not gotten any complaints about fairness from parents. This acceptable/not acceptable practice has made grading a lot quicker and easier for me. I do not labor over borderline decisions about whether a project merits a "B-" or a "C" grade. I no longer feel bad thinking that Joe made a simple error and now it is too late for him to earn a higher grade or that Maria shows little understanding of an idea and it is too late for her to revisit it because the projects are all submitted. It is not too late for either of them. Joe can still make corrections after he knows what needs to be fixed and he fixes it. Maria can still get a tutorial from me or a classmate and further develop her understanding. In other words, everyone still has the opportunity to learn and meet the goals.

My colleagues in CAPITAL first started this assessment practice and I must admit that, in the beginning, I was reluctant to grade projects on an acceptable / not acceptable basis. I felt that it would not be fair to those students who completed their work accurately and on time to allow other students who needed to make corrections to earn the same grade. In our group discussions about this practice, the focus of the assessment of projects was also shifting more toward the demonstration of the concepts and skills and less on the aesthetic quality of the work. I had difficulty justifying giving a student who turned in an elaborate project that demonstrated much effort and time spent the same grade as a student who turned in a project that appeared to be done quickly with less attention to the overall appearance of the work. I feared that this new assessment approach would not provide the incentive to students to do the high quality work that some of my high achieving students produced, knowing that their grades would be based solely on the scientific concepts and skills and less on appearance.

I have gotten some of the biggest rewards of my teaching career by seeing the looks of pride and smiles that come over the faces of students who had revised time after time and finally got it!

Now I look at it differently. I began to ask myself what it was that I really valued when I assigned and evaluated student work. I have come to the conclusion that, for me, the value lies in what the students have learned. It does not matter that some students take longer and others are quick learners or some are better at aesthetically expressing themselves than others. Of course, I still require my students to turn their work in on time and since I have instituted these revision opportunities into my practice, the on-time turn in rate has been very high—even for the students who do not regularly turn in their work. I think this was due to the fact that the students knew they had to have their work prepared for the due date in order to be eligible for the revision time. They seem more willing to revise their work knowing that the possibility of earning an "A" for demonstrating their understanding is still there. The main thing is they learn the science concepts and skills we are focused on in class and this is what I now think of as the center of my assessment practices. If they demonstrate their understanding or abilities at any time, they have met the goals of the class and earned that "A". It took me a while to accept that last statement. But through lots of conversations with my colleagues and slowly trying new ideas, I am now finding that this is working for my students and for me. I even heard myself suggesting this assessment approach to a new teacher at an assessment workshop we recently held in our school.

Changing Mindsets About Classroom Assessment

For all five teachers, the general shift in thinking about assessment was toward making learning goals and expectations more clear to the students and shifting from an emphasis on grading student work after it is completed and submitted to an emphasis on providing immediate feedback that supports the students while working toward the learning goals. Vicki summarized this shift: "the most valuable thing we can do to help with our students' learning is to really *be there* in class, giving as much feedback as we can directly to the students to support them as they engage in the learning of new concepts and skills."

All of the teachers instituted the strategy of assessing projects using an "acceptable / unacceptable" format as described by Elaine. As this approach to assessing projects spread throughout the group, each teacher made modifications to tailor the approach to his or her own practice. In the group conversations, the teachers described this assessment approach as one in which feedback to the students was more consistent because the standards and criteria for the work were more clear and focused on the essential conceptual understandings expected of students. Joni strongly felt that the revision process afforded by the acceptable / not acceptable assessment strategy provided a more equitable opportunity for learning for all of her students:

> I have leveled the learning field. At my school, I have many students who have special learning needs, students who might not get it the first or second time. This process allows those students to keep trying until I know they understand the concepts better. I have gotten some of the biggest rewards of my teaching career by seeing the looks of pride and smiles that come over the faces of students who had revised time after time and finally got it! The biggest benefit is that I have few students who do not complete projects. Since the emphasis is on learning and the goal is an "A" for accepted, everyone has an equal opportunity to achieve success.

From the discussions about assessment strategies that are focused on learning, the teachers' mindsets about their role as the teacher shifted. The conversations among the teachers began with a fundamental desire to reduce the amount of paperwork they felt they had to maintain to be fair and consistent in calculating grades for students. The teachers felt trapped in a cycle of collecting homework and projects, grading assignments with the primary goal of recording a score in their grade book, and returning the marked papers to the students in a timely manner (which, given that each of the teachers is in contact with approximately 180 students every day, this was not always possible). As

As we open the door for students to turn a failure into a success, we must allow our peers a comfortable environment to do the same.

the teachers' conversations turned to discussing ways to make expectations clearer to students and their appreciation grew for what information they could glean about a student's understanding through daily interactions, some of the teachers began to play a different role in their classrooms. The management of paper and grade spreadsheets is a responsibility, but the amount of time spent on these activities has been greatly reduced in favor of time spent designing meaningful projects with clear standards, daily interactions with students in both one-on-one and whole-class discussions centered on formative assessment, and on creating more opportunities for peer feedback and self assessment.

This group has also engaged their colleagues in conversation about what they have learned in the examination of their own practices. These efforts to influence change in others, though, present their own challenges, as described by Neil below.

Challenges in Changing Practices: Neil's Experience

My CAPITAL peers and I have been trying to figure out how to address some of the assessment issues we have learned about in our own classrooms with other teachers in our district in a way that does not sound like we are suggesting that they have been doing it wrong all these years. Many lunchtime conversations have proven how difficult it is to tread these waters. The quick wall of, "No, that won't work for me" often pops up at any mention of the successes we have experienced in our own classrooms. One teacher we work with is very caring and concerned about his students. He gets very involved in the success of his students as shown in their grades. When his students fail, he looks for interventions for the students, but does not look at his own teaching or assessment practices. He searches for the difficulty the student might have as an individual or at home, but he does not view his role as actively guiding the student toward successful learning. When planning together, I am often confronted with his resistance to thinking about how formative assessment might make a difference in what he understands about his students' performance and understanding.

After a year and half of talking about assessment related issues and sharing some of my new practices during our ongoing planning time together, I have seen a slow, steady progression moving him from his teacher-centered approach to his starting to look at what the students are learning. He saw our excitement and results, and began to feel his way into the processes that got us there. A year ago, he was not ready to adopt completely a new way of thinking about assessment as it relates to student learning. Now, there is a willingness to talk about new ideas and tailor some for his own classroom.

In contrast, during one of our many lunchtime conversations about out teaching another teacher became interested in our group's conversation about the acceptable / unacceptable assessment practices. She asked some questions about how it worked and the advantages that we saw. That was all it took. A week later, she was sharing her own students' successes with the process. She went from conversation to successful implementation without any further prodding or discussion. She saw the benefits and fit it into her practice without any hesitation.

One of the ways that teachers have to make sense of their everyday activities in their classrooms is to talk with their colleagues about what is going on.

Why was the second teacher so easy to convince while the others are so difficult? As I ponder this I am forced to believe that classroom practices are the adult equivalent of our students' work. They need to be open for revision and teachers need to know that it is okay to admit that their work needs revision. As we open the door for students to turn a failure into a success, we must allow our peers a comfortable environment to do the same. I think the key to this is realizing that as our kids have individual learning needs, teachers do too. We all have buttons we do not like pushed and closets we want to keep closed. We need to foster an environment at the school sites where trying new practices in the classroom is acceptable, open for discussion, and an expected way for teachers to get better at what they do. We also need to provide opportunities for teachers to collaborate with each other on a regular basis. One of the ways that teachers have to make sense of their everyday activities in their classrooms is to talk with their colleagues about what is going on. We found the collaboration part of our work with CAPITAL to be an essential part of our learning.

CAPITAL: A Collaborative Approach to Changing Classroom Practice

As stated earlier, CAPITAL was guided by the theoretical perspectives of practical reasoning with the teacher as the primary driver of the choice of action and decision in the classroom. Deliberation is the essential process of practical reasoning, requiring the person who acts to size up the situation and view it with both the desired outcome and the appropriate means toward that outcome in mind (Schwab, 1969). The collaborative process of sharing ideas, receiving feedback from colleagues, and reflecting on practice and underlying beliefs with peers was a central design feature of CAPITAL that enabled the teachers to deliberate with their peers as they engaged in the process of making changes in their daily classroom practices. CAPITAL staff intended that the group conversations would provide opportunities for the teachers to learn not only from their own practice, but also from the practice of their colleagues (Wenger, 1998).

The teachers from New Haven felt strongly that the collaboration in which they engaged was a central feature of their change process. However, they point out that collaboration is not something that they knew how "to do" simply by virtue of being a teacher. When given the time and opportunity to work with their peers, the teachers pointed out what is often left unasked in teacher collaborations.

> Sharing ideas and building lessons together is really great, but no one ever asks, "How do we know this lesson will be effective?" In fact, the very question would probably be seen as insulting. How dare we question each other's effectiveness as teachers? The question would probably have been met with, "We'll find out when we give the test, I guess."

The group pointed to the importance of having a facilitator for their initial collaborative conversations for several purposes. The facilitator kept the group focused and on track and provided a process for the discussions. More importantly, from the group's perspective, the facilitator asked probing questions and encouraged the teachers to see these questions as models for questions they could ask one another. The group also described how the facilitator helped them pause on issues that they might have left unexamined.

> For example, someone might say, while discussing a lesson, "Of course, I can't talk to every student about their grade." We, the teachers, probably would have let the statement go, waiting instead to hear more about the lesson itself. One of the facilitator's jobs is to listen for such statements and revisit them. So, the facilitator might say, "You said you can't talk to every student about his or her grade. How do you decide who to talk to? How does everyone else know about their grades?" Some of our most interesting discussions came out of statements that we revisited through probing by the facilitator.

With the opportunity to not only deeply think about their practice and to talk about it with trusted colleagues,

When given the time and opportunity to work with their peers, the teachers pointed out what is often left unasked in teacher collaborations.

this group asked each other those typically unasked questions. They became more comfortable and adept at probing each other on the underlying reasons for their practical choices.

CAPITAL set out to help teachers bring about changes in their everyday assessment practices and to learn about everyday assessment from these teachers' classroom work. The fundamental shifts in understanding the purposes of assessment expressed by the teachers were catalyzed by actions they took to change their practice and those actions further informed the teachers' beliefs about the purpose and use of assessment information. The collaborative process of exchanging ideas, seeking reasons for actions, and exploring alternative strategies helped these teachers make sense of their change process and further clarify their own underlying beliefs and assumptions about the purpose of assessment in their classrooms.

References

Atkin, J. M., J. E. Coffey, S. Moorthy, M. Sato, and M. Thibeault. (2005). *Designing everyday assessment in the science classroom*. New York: Teachers College Press.

Black, P. J., and D. Wiliam. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*(4), 474–482.

Coffey, J. E., M. Sato, and M. Thibeault. (2005). Classroom assessment up close—and personal. *Teacher Development, 9*(2), 169-184.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438–481.

Duschl, R. A., and D.H. Gitomer. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*(1), 37–73.

Gauthier, D. P. (1963). *Practical reasoning: The structure and foundations of prudential and moral arguments and their exemplification in discourse*. Oxford, England: Oxford University Press.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded version*. J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Washington, DC: National Academy Press.

Palincsar, A. S., and A. L. Brown. (1984). Reciprocal teaching of comprehension monitoring activities. *Cognition and Instruction, 1*(2), 117–175.

Sato, M., J. E. Coffey, and S. Moorthy. (2005). Two teachers making assessment for learning their own. *Curriculum Journal, 16*(2), 177-192.

Scardamalia, M., C. Bereiter, and R. Steinbach. (1984). Teachability of reflective processes in written composition. *Cognitive Science, 8*: 173–190.

Schon, D. A. (1983). *The reflective practitioner*. U.S.A.: Basic Books.

Schwab, J. J. (1969, November). The practical: A language for curriculum. *School Review*, 1-23.

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Education Research Journal, 33*(2), 359–382.

Wenger, E. (1998). *Communities of Practice*. Cambridge: Cambridge University Press.

White, B. Y., and J.R. Frederiksen. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Vicki Baker teaches 7th and 8th grade science at Alvarado Middle School in Union City, California.

Elaine Fong teaches 8th grade science at Alvarado Middle School in Union City, California.

Joni Gilbertson teaches 7th grade science at Barnard White Middle School in Union City, California.

Tracey Liebig teaches 7th grade science at Alvarado Middle School in Union City, California.

Neil Schwartzfarb teaches 6th grade math and science at Alvarado Middle School in Union City, California.

Mistilina Sato is assistant professor of Teacher Education and Science Education at the University of Minnesota. As a graduate student and post-doctoral fellow at Stanford University, her research and professional development work with teachers focused on teacher leadership, National Board Certification for teachers, performance assessment for beginning teachers, formative assessment in science classrooms, and elementary science education reform. Correspondence concerning this article may be sent to msato@umn.edu

Teacher Research: Exploring Student Thinking and Learning

Teachers who conduct research in the context of their own teaching practices can contribute to knowledge about reform-based instruction.

Introduction

What does it mean for teachers to do research? Is forming a fleeting question in the midst of teaching an act of research? Making copies of student work and discussing these with colleagues? Videotaping small group activities and writing about what students said and did? Analyzing performance differences among classes taught in different ways? Presenting or publishing findings about student thinking and learning?

All of the above seem to me to be acts of research to be celebrated and nurtured (van Zee, 2005; van Zee & Roberts, in press). As an organizer of Teacher Researcher Day during national conferences of the National Science Teachers Association (http://www.nsta.org), I welcome a wide spectrum of participation, from teachers just beginning to ask questions about their students' learning to presenters who have published their own studies. Some authors, however, prefer to reserve the word "research" for more formal investigations (Hammer and Schifter, 2001; Richardson 1994).

Several phrases refer to research conducted by individuals in the context of their own practices. For example, three special interest groups (SIGs) of the American Educational Research Association (http://www.aera.net) focus on such research. These include the Teacher as Researcher SIG, which "is dedicated to supporting research done in schools by PreK-12 practitioners on their own practice" (http://www.teacherasresearcher.org). Participants in the Self-Study of Teacher Education Practices SIG are "teacher educators who are working on the problems of education through the study of their own practices" (http://www.ku.edu/%7Esstep/about.htm). The purpose of the Action Research SIG is "to involve teacher-as-researchers, administrators, researchers, and community members in collaborative action research that examines educational practice and encourages educational reform and improvement" (http://explorers.tsuniv.edu/ar-sig/).

The potential of including teachers in the research enterprise has been recognized for many decades.

The emerging importance of teacher research also is evident from publications, including books about methods of conducting such research (Cochran-Smith & Lytle, 1993; Hopkins, 2002; Hubbard & Power, 1993, 1999; McNiff & Whitehead, 2005; Mills, 2003). A chapter on science teachers as researchers will appear in the upcoming *Handbook of Research on Science Education* (Roth, in press). For the first time, a chapter on practitioner research was included in the fourth edition of the *Handbook of Research on Teaching* (Richardson, 2001). This chapter provided a detailed review of the history and various forms of research in which individuals have investigated their own practices in the context of whatever roles they have played in educational settings (Zeichner & Noffke, 2001). A bibliography of publications about the scholarship of teaching and learning in higher education (Hutchings & Bjork, 2002) is available at http://www.carnegiefoundation.org/elibrary/docs/bibliography.htm.

The potential of including teachers in the research enterprise has been recognized for many decades. Early in the 20th century, for example, John Dewey (1928/56) envisioned teachers as producing "a series of constantly multiplying careful reports on conditions which experience has shown in actual cases to be favorable or unfavorable to learning" (p. 125-126). Late in the 20th century, Eleanor Duckworth (1989) articulated a similar vision in her essay "Teaching as Research." She proposed that teachers could legitimately "contribute to the theoretical and pedagogical discussions on the nature and development of human learning" (1987, p. 168). Cochran-Smith & Lytle (1990, 1999) reflected on many of the issues that arise, however, when research on teaching is conducted by university researchers observing others teach or by teachers on their own practices.

In the *National Science Education Standards*, the National Research Council (1996) recommended that "Professional development activities must … provide opportunities to learn and use the skills of research to generate new knowledge about science and the teaching and learning of science" (p. 68). In a subsequent publication, the Council advocated models of professional development that include reflection, intellectual engagement, and advancement in the profession (NRC, 2001, p. 103). Feiman-Nemser (2001) described a "new paradigm" for professional development in which "teachers must be able to ask hard questions of themselves and their colleagues, to try something out and study what happens, to seek evidence of student learning, and explore alternative perspectives" (p. 1040). Such studies are a form of teacher research and can evolve into systematic efforts that generate knowledge that other teachers find useful.

Most teacher research yields detailed descriptions and interpretations specific to particular situations. Such reports are typical of qualitative research rather than quantitative studies. Differences between these research paradigms have been described in terms of several criteria (Donmoyer, 2001). The emphasis in qualitative research is on transferability (whether a reader views findings as applicable to the reader's own situation) rather than generalizability (whether the findings apply anywhere at any time). Also considered is whether enough sources were consulted over a long enough time period in thorough enough ways (trustworthiness) rather than whether others can get the same results (reliability). A third criterion is whether the interpretations make sense (credibility) rather than whether what was measured matches well with what was asked (validity).

Most teacher research yields detailed descriptions and interpretations specific to particular situations.

Teachers who conduct research in the context of their own teaching practices can contribute to knowledge about reform-based instruction. Examples below include documenting inquiry-based science instruction, developing documentary web sites to share findings about science learning and teaching, contributing to knowledge generation at state and national as well as local levels, addressing the needs of at-risk students, and forming collaborative research communities. These examples illustrate research by elementary, middle school, high school, and college faculty as well as the use of data such as transcripts of discussions, reflective journals, and copies of student work. The paper closes with some suggestions for getting started with teacher research.

Documenting Inquiry-based Science Instruction

Teachers who document their students' learning can help interested colleagues in envisioning the inquiry-based science instruction recommended in the National Science Education Standards (National Research Council, 1996). In Doing What Scientists Do: Children Learn to Investigate Their World, for example, Ellen Doris (1991) describes ways in which she engages children in making observations and generating questions that they can explore.

In discussing her role as a facilitator of learning, Doris presents a transcript of seven- and eight-year old children sharing their observations of crickets with one another (Doris, 1991, p. 91): See Figure 1.

The transcript continues with the children's suggestions for what they might do to find out about cricket eyes. In writing about these students' observations and questions, Doris reflects upon both the students' thinking and her role as the teacher in eliciting that thinking. She notes the importance of extending the initial observations children make, of calling attention to differences in their observations, of recognizing assumptions being made, and of helping children figure out how to go about answering their own questions.

Other books by elementary teachers documenting their own teaching practices include *Nurturing Inquiry: Real Science for the Elementary Classroom* by Charles Pearce (1999), *Talking Their Way into Science; Hearing Children's Questions and Theories, Responding with Curricula* by Karen Gallas (1995) and *Science Workshop: Reading, Writing, and Thinking like a Scientist* edited by Wendy Saul, Jean Reardon, Charles Pearce, Donna Dieckman, and Donna Neutze (2002).

Figure 1

Marty:	They have long feelers that look as if they come out of their eyes.
Beth:	I noticed that my cricket has two eyes.
Jeff:	Some have two little tails on the back and some have the two little tails with one big tail in the middle.
Jackie:	My cricket has four eyes.
Teacher:	Wait a minute! When Beth looked at her cricket, she found two eyes. When Jackie looked, she found four! Did anyone else see eyes on a cricket?
Nicky:	Mine had two.
Maureen:	Mine too!
Lynnie:	I think mine had four.
Jeff:	Mine definitely had two.
Greg:	I think that they have lots and lots of eyes—they have a special kind of eyes.
Teacher:	Compound eyes?
Greg:	That's it.
Teacher:	Claudia, you look puzzled.
Claudia:	But how do you know where the eyes are? I didn't see <u>any</u> eyes!
Douglas:	They're on the head—one on each side—right where the feelers are.
Claudia:	But how can you tell those are eyes? Just because our eyes are on our heads doesn't mean a cricket's are!

Developing Documentary Web sites to Share Findings about Science Learning and Teaching

Teacher research also includes developing documentary web sites to share one's findings about science learning and teaching. A middle school science teacher, Claire Bove, for example, created a web site to show how she establishes a community within her classroom, encourages a playful approach to experimentation, engages students in developing explanations and questions as well as in talking and arguing about science (http://feelingathome.org).

Developed with support from the Carnegie Foundation for the Advancement of Teaching, Bove's web site includes examples of students' writings and drawings with commentary by the teacher as well as video clips of students in action. Also included are video clips of the author and a colleague talking together about their teaching practices.

Teachers who conduct research in the context of their own teaching practices can contribute to knowledge about reform-based instruction.

In a reflection comparing science discourse in her middle school classroom and the cell biology research lab in which she had worked as an undergraduate, Bove wrote:

> One of the most interesting parts of the [undergraduate research assistant] job was going to the weekly lab meetings … The main difference between the talk at these lab meetings and the talk in a science classroom was that no one in the lab knew the answers to the questions we were asking … It is hard to have this kind of genuine discussion in a middle school classroom …
>
> Because the teacher does know the answer, students are used to trying to figure out what the teacher thinks the answer is. To change this classroom dynamic, and to get students to try and

> figure out what they themselves think the answer is, I try to frame classroom discussions so that I ask about students' experience of, and opinions about, the things we are studying.
>
> For example, to bring in their own experience, I ask them if they have ever felt an earthquake, and what happened when they did … And to encourage them to state their own opinions, I frame questions like this: "Jerrad said that a pinto bean will sprout if you get it wet, but a popcorn kernel won't. Do you agree or disagree?" When the question implies that it is another student who is proposing a science idea, instead of the teacher, it is much easier for a student to put forth his or her own idea.
>
> Before we have the class discussion, I often ask students to find out what the person next to them wrote. And then I ask them to tell what they, or the person next to them, answered.
>
> All of these techniques are attempts to help students practice saying what they think, and to defend what they think, and often to change what they think. In short, to practice discourse as a way to learn to think.

Under construction is an addition to the web site with a detailed presentation of student explorations about density (Bove, 2006).

Such documentary web sites make available examples of inquiry-based instruction, particularly for prospective teachers, who may not experience reform approaches in their placement classrooms. A teacher educator, Anna Richert at Mills College, for example, uses Bove's web site, as well as others, as "texts" in her course on adolescent development. She has documented this use on her own web site (see http://quest.carnegiefoundation.org/~arichert/section2A.html). In a web page entitled "Getting Started: Learning to Learn from the Practice of Others," she reports on her invitation to the prospective teachers in her class to investigate the following questions: "What do these teachers know about their learners? How do they come to know their learners? And how do they use this knowledge in their teaching?" On a web page entitled "Studying Claire's Site," Richert provides links to a lesson plan, study guide, prompt page, and syllabus as well as video clips of prospective teachers discussing Claire's web site. By perusing Richart's web site, interested teacher educators can learn from her experiences in using the K12 documentary web sites to foster learning about teaching.

Known as "snapshots of practice," other web sites developed with Carnegie Foundation support are collected in an online "gallery" (see http://gallery.carnegiefoundation.org). Several of these are discussed in detail in an anthology (Hatch et al, 2005). They represent forms of making public these teachers' "wisdom of practice" (Shulman, 2005).

Contributing to Knowledge Generation at State and National as well as Local Levels

Some teacher researchers contribute to knowledge about teaching and learning not only by conducting workshops in their schools and districts but also by presenting at conferences, publishing their findings, and applying their results at state and national levels. A high school physics teacher, Jim Minstrell, for example, began documenting his students' learning by asking his students diagnostic questions at the beginning of a unit and shaping instruction according to their responses (1982). Funding from government agencies and private foundations enabled Minstrell to devote some of each day to his research as well as to teaching. He viewed his classroom as a research site for studying physics learning and teaching by himself, his students, two mathematics teachers whom he was coaching to teach physics, and university researchers whom he invited to participate in his projects.

Over many years, Minstrell and his colleagues developed a framework of Facets for Thinking that is aligned with Washington State's Essential

Some teacher researchers contribute to knowledge about teaching and learning not only by conducting workshops in their schools and districts but also by presenting at conferences, publishing their findings, and applying their results at state and national levels.

Academic Learning Requirements and the *National Science Education Standards* (National Research Council, 1996) (see http://www.facetinnovations.com/main/facet-think.htm for middle school science and mathematics and high school physics). For example, the *National Science Education Standards* includes the following content standard for Grades 5-8: "The motion of an object can be described by its position, direction of motion, and speed. That motion can be measured and represented on a graph." (NSES p154, grades 5-8). The cluster of facets related to motion includes several ways in which students view a graph:

> **90** Student views a position or speed graph as a map of the actual motion.
> **91** Student interprets an upward (or downward) sloping graph to mean the object is going up hill (or downhill).
> **92** Student interprets a flat line on the graph to mean the object is moving on a flat surface.

The framework of facets of student thinking has served as the basis for building a diagnostic assessment system for promoting learning (Minstrell, 2000, 2001) and development of a computer program, the Diagnoser, that students can use to check their understanding (see http://www.diagnoser.com/diagnoser/).

Focusing on the Needs of "At Risk" Students

College instructors also have been documenting their own teaching practices. A professor of chemistry, Denis Jacobs (2000), for example, developed a web site to share with other faculty ways in which he redesigned a general chemistry course to address the needs of "at risk" students (see http://gallery.carnegiefoundation.org/djacobs/). The web site presents his rationale, ways he implemented cooperative learning strategies, and assessments of the impact of these reforms on students' performance, attitudes, and subsequent success in advanced science courses. One page of the web site compares the traditional and alternative sections of a course in terms of numbers of students enrolled in lectures (about 250 in each), recitations (optional for the traditional section, mandatory for the alternative section), and laboratories (mixed enrollment from both sections). Jacobs describes what he does during lectures as follows:

> Students pair off to discuss conceptual questions in lecture. Periodically during each lecture, I ask a conceptual question to the 250 students in the class. The students vote, through a show of hands on various responses to the question. Then I ask the students to turn to someone sitting next to them and to explain their reasoning. After two or three minutes of one-on-one discussion, the class votes again on the possible answers. Next, I ask representative students who voted on different responses to share their reasoning with the entire class. We often perform a live experiment to test which response is most correct.
>
> Eric Mazur, a physicist from Harvard University, developed this approach to peer instruction and called it 'concept tests.' Faculty at the University of Wisconsin and Carnegie Mellon University have written similar 'concept test' questions for general chemistry.

By providing links to web sites by colleagues at other universities, Jacobs fosters communication of reform practices that he has found useful in the context of the large lecture courses typical of science instruction at major universities. The web site also includes video clips of cooperative learning during recitation sections where students work in small groups to solve challenging problems. In addition, the web site documents dramatic improvement for "at risk" students in terms of their performance on general chemistry exams, performance in related science courses, choice of majors, and attitudes.

Forming Collaborative Research Communities

Teachers in a school may collaborate with one another and/or with university researchers on long-term research projects. In *Creating Scientific Communities in the Elementary Classroom*, for example, Maureen Reddy, Patty Jacobs, Caryn McCrohon, and Leslie Rupert Herrenkohl (1998) report upon their collaboration as teacher and university researchers. Their book includes sections written by the teachers as well as by the university researchers in interpreting what was happening during implementation of a new science curriculum in two second grade classrooms. Caryn McCrohon, for example, reflected upon her disappointment with the curriculum's worksheets and subsequent shift to using dialogic journals:

For both research scientists and teacher researchers, acts of research begin with curiosity about phenomena observed.

> For the most part the worksheets were coming in filled out, but I quickly discovered that they did not help me assess what the children were learning. The worksheets required the children to record results, or answer one or two factual questions. They did not ask the type of questions that stimu-

lated descriptions of procedures, explanations, or even children's general thinking about the activities …

With advice from her principal, she shifted to having the students write journals:

> Once we were finished with our wrap-up my students had about thirty minutes to write a journal entry and draw a picture to go with the entry. At first I was a bit wary. Science was taking longer than expected out of our seven-hour day. My principal said not to worry because science journals are part of language arts and she was excited to see writing across the curriculum. (p.94).

Following Caryn's reflections is a section written by the university researchers that provides examples of many kinds of journal entries by the children and their teacher. The collaborative construction of this account of the use of dialogic journals makes available to interested teachers both the perspective of the classroom teacher and the detailed analysis developed by university researchers. Other accounts of collaborative research communities include *Changing Schools From Within: Creating Communities of Inquiry* by Gorden Wells (1993) and *Teacher Research for Better Schools* by Marion Mohr, Courtney Rogers, Betsy Sanford and Marion Nocerino (2002).

Getting Started with Teacher Research

For both research scientists and teacher researchers, acts of research begin with curiosity about phenomena observed. A good way for teachers to get started is to listen closely to what their students say, to watch what their students do, and to record any questions that such observations prompt them to wonder. By writing reflective journals, teachers can begin documenting what is happening in their classrooms.

Interested teachers should consult with their principals about the need for district approval of more formal studies. Although obtaining permission from parents involves some effort, placing an audio- and/or video- camera unobtrusively in the classroom and taping instruction can capture interesting conversations about science for later reflection and analysis. Making copies of a variety of student work can trace students' growth and document various ideas they generate.

School administrators can foster teacher research by providing equipment, supplies, journals, and most importantly, time during the school day for teacher researchers to meet regularly to discuss their studies with one another.

Teacher researchers can learn from one another by meeting regularly to discuss video clips of their students in action or a set of student work. Taking the next step to share findings through school and/or district workshops can deepen one's own understandings as well as communicate insights and experiences. Presenting at conferences and writing for publication can inspire others to begin improving their practices through such inquiries.

School administrators can foster teacher research by providing equipment, supplies, journals, and most importantly, time during the school day for teacher researchers to meet regularly to discuss their studies with one another. District and state administrators can assist by designating resources for teacher research and by sponsoring teacher researcher workshops and meetings. In addition, funding for travel to present at conferences can help interested teachers become part of a broader research community. County funding, for example, made possible participation in a NSTA national conference of a teacher researcher who described his experiences there as follows:

> Teacher Researcher Day allowed me to stand in one room (literally), look around, and say, 'Wow! Here's a group of individuals dedicated to teaching/learning. It doesn't get any better than this!' I felt as though it elevated what I do as an individual and lends credibility to me, the classroom teacher. We had a ballroom full of excited individuals talking about student thinking and what that thinking meant. [teacher researcher, NSTA national conference, April 2003]

Through experiences such as these, teachers can learn from one another how to formulate questions about science teaching and learning, collect data as they teach, develop interpretations of these data, and share their findings with others. Such classroom-based research has the potential to inspire teachers to improve instruction through learning about, trying out, and adapting practices advocated in reform documents.

References

Bove, C. (2006, April). What happens when even the teacher doesn't know what the next experiment will be? Annual meeting of the American Educational Research Association, San Fra ncisco.

Cochran-Smith, M. and S.L. Lytle. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher,* 19 (2): 2-11.

Cochran-Smith, M., and S. L. Lytle. (1993). *Inside/outside: Teacher research and knowledge*. New York: Teachers College Press.

Cochran-Smith, M. and S. L. Lytle. (1999). The tea cher research movement: A decade later. *Educational Researcher*, 28(7), 15-25.

Dewey, J. (1928/1956). Progressive education and the science of education. In M. S. Dworkin (Ed.) *Dewey on education selections* (pp. 113-126). New York: Teachers College Press.

Donmoyer, R. (2001). Paradigm talk reconsidered. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 174-197). Washington, D.C.: American Educational Research Association.

Doris, E. (1991). *Doing what scientists do: Children learn to investigate their world*. Portsmouth, NH: Heinemann.

Duckworth, E. (1987). *The having of wonderful ideas and other essays on teaching and learning*. New York: Teachers College Press.

Feiman-Nemser, S.(2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, 103(6), 1013-1055.

Gallas, K. (1995). *Talking their way into science: Hearing children's questions and theories responding with curricula*. New York: Teachers College Press.

Hammer, D., and D. Schifter. (2001). Practices of inquiry in teaching and research. *Cognition and Instruction*, 19(4), 441-78.

Hatch, T., D. Ahmed., A. Lieberman., D. Faigenbaum., M.E. White., and D.H. P. Mace, (Eds.) (2005). *Going public with our teaching: An anthology of practice*. New York: Teachers College Press

Hopkins, D. (2002). *A teacher's guide to classroom research*. Maidenhead, Berkshire, UK: Open University Press.

Hubbard, R. S. and B.M. Power. (1993). *The art of classroom inquiry: A handbook for teacher research*. Portsmouth, NH: Heinemann.

Hubbard, R.S. and B.M. Power. (1999). *Living the questions: A guide for teacher-researchers*. York, Maine: Stenhouse Publishers.

Hutchings, P. and C. Bjork. (2002). The scholarship of teaching and learning in higher education: An annotated bibliography. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Jacobs, D. C. (2000). A chemical mixture of methods. In P. Hutchings (Ed.), *Opening lines: Approaches to the scholarship of teaching and learning* (pp. 41-52). Palo Alto, CA: Carnegie Foundation.

McNiff, J. and J. Whitehead. (2005). *Action research for teachers: A practical guide*. London: David Fulton Publishers.

Mills, G. (2003). *Action research: A guide for the teacher researcher*. (2nd ed.) New Jersey: Merrill Press. (http://www.sou.edu/education/action_research.htm)

Minstrell, J. (1982). Explaining the "at rest" condition of an object. *The Physics Teacher*, 20, 10-14.

Minstrell, J. (2000). Student thinking and related assessment: Creating a facet assessment-based learning environment. In J. Pellegrino, L. Jones, and K. Mitchell, (Ed.), *Grading the nation's report card: Research from the evaluation of NAEP*. Washington DC: National Academy Press.

Minstrell, J. (2001). The role of the teacher in making sense of classroom experiences and effecting better learning. In D. Klahr, S. Carver, (Ed.), *Cognition and Instruction: 25 years of progress*. Mahwah: Lawrence Erlbaum Associates.

Mohr, M., C. Rogers., B. Sanford., and M.A. Nocerino. (2003). *Teacher research for better schools*. New York: Teachers College Press.

National Research Council. (1996). *National science education standards*. Washington, D.C.: National Academy Press.

National Research Council (2000). *Educating teachers of science, mathematics, and technology: New practices for the new millennium* (Committee on Science and Mathematics Teacher Preparation). Washington, DC: National Academy Press.

Pearce, C. R. (1999). *Nurturing inquiry: Real science for the elementary classroom*. Portsmouth, N.H.: Heinemann.

Reddy, M., P. Jacobs., C. McCrohon., and L.R. Herrenkohl. (1998). *Creating scientific communities in the elementary classroom*. Portsmouth, NH: Heinemann.

Richardson, V. (Ed.) (1991). *Handbook of research on teaching*. Washington, D.C.: American Educational Research Association.

Richardson, V. (1994). Conducting research on practice. *Educational Researcher*, 23(5), 5-10.

Roth, K. (in press). Science teachers as researchers. In S.K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education*, Mahwah, NJ: Erlbaum.

Saul, W., J. Reardon, C. Pearce, D. Dieckman, and D. Neutze. (2002). *Science workshop: Reading, writing, and thinking like a scientist*. Portsmouth, NH: Heinemann.

Shulman, L.S. (2004). *The wisdom of practice: Essays on teaching, learning, and learning to teach*. San Francisco: Jossey-Bass

van Zee, E.H. (2005). Teaching "science teaching" through inquiry. In K. Appleton (Ed.), *Elementary science teacher education: Contemporary issues and practice* (pp. 239-257). Mahwah, NJ: Lawrence Erlbaum.

van Zee, E. and D. Roberts. (2005). Engaging teachers in research on science learning and teaching. In J. Rhoton and P. Shane (Eds.), *Teaching Science in the 21st Century*. Arlington, VA: National Science Teachers Association Press.

Wells, G. (1993). *Changing schools from within: Creating communities of inquiry*. Portsmouth, NH: Heinemann.

Zeichner, K. M. and S.E. Noffke. (2001). Practitioner research. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 298-330). Washington, D.C.: American Educational Research Association.

Emily van Zee is associate professor of science education at Oregon State University, Corvallis, Oregon. Correspondence concerning this article may be sent to Emily.vanZee@science.oregonstate.edu

John H. Holloway

Connecting Professional Development To Student Learning Gains

An argument is made that highly effective, research-based professional development can translate into improved student achievement and overall school effectiveness.

Introduction

Is there a direct link between teacher professional development and an increase in student achievement? There seem to be two answers to this rather obvious question: "Maybe" or "It depends". "Maybe" if we can find an accurate way to measure the impact of professional development. And, "It Depends" if the professional development chosen and the school climate are both aligned with our goals. Over the years, teachers and school leaders have engaged in some form of training beyond their formal preservice, university preparation programs. Whether referred to as in-service training, professional day, or staff or professional development this activity normally had one purpose: to improve participant's knowledge or skills. But, if this improvement in teacher skills actually occurred, did it necessarily translate into increased student learning?

My earliest experiences with professional development, at the start of my career over 40 years ago, as a new high school science teacher may not be so different from what many teachers experience today. During my first year teaching all teachers attended a mandatory "Professional Day". At this morning session we all filed into the school auditorium for a presentation by a local psychiatrist. I never really remembered what he talked about, but the session was memorable, nevertheless. In the middle of his presentation a worker walked onto the stage, next to speaker's podium, and began to tune the school piano to the great embarrassment of the superintendent. During my second year of teaching our district "Professional Day" had a similar, mandatory attendance format for all teachers, regardless of subject or grade level taught. This session was presented by Professor Hubert Alyea of Princeton, NJ. Alyea was, and still is, known for his fantastic classroom demonstrations that made chemistry come alive. His session was highlighted by a demonstration of "rates of reaction" in which he prepared several large flasks of different, colorless solutions. He sang a song with lyrics that included all the Ivy League universities. The lyrics were timed to coincide with the change of color of the solutions to match the university color he was singing about. I'm pretty sure the first experience had no impact on my teaching. The second was certainly motivating and gave me ideas for my own lesson planning, but did it translate into increased student learning? What effect that second experience had on other faculty members was unknown. With an ever increasing demand for accountability for student achievement, how can we increase the chances that professional development will be transformed into increased student improvement? To answer this question we must examine three, more specific issues: Does professional development impact student learning? How we can judge the effectiveness of our professional development program? And, what professional development characteristics are linked to improved student learning?

With limited teacher time to attend in-service training and limited financial resources to pay for this training, districts are searching for training that provides returns on these investments in time and money: returns in terms of increased student achievement.

Does Professional Development Influence Student Achievement?

Unfortunately there are still too many districts that see professional development as the one shot "Professional Day" of my earliest experiences. With limited teacher time to attend in-service training and limited financial resources to pay for this training, districts are searching for training that provides returns on these investments in time and money: returns in terms of increased student achievement.

Barry Fishman and his colleagues (April 29, 2000) maintain that professional development is all about a process of giving teachers new skills and concepts related to the work of teaching. Beyond this acquisition of new skills, however, is an assumption that this will be translated into improved student learning and achievement once the teacher applies these new skills or knowledge to practice in the classroom.

Not all professional development experiences are effective in bringing about student learning gains however, but studies are beginning to emerge that show that those that are research-based, thoughtfully conceived and delivered, and focused on the right things can, indeed, impact learning. Thomas Guskey (Winter 2005) has had considerable experience in examining the effect of professional development on student learning. He has found that powerful professional development will help the educator acquire the instructional procedures and scientifically researched-based strategies they need to help all students reach the articulated learning goals. It is important to focus on improving the teacher, according to Guskey, because true educational reform does not take place at the state or district level. Unless, it occurs at the school building and classroom levels, student improvement is unlikely because improvement in education is defined by more students learning better, and that only occurs in the classroom.

Susan Murphy (Spring 2005), like Guskey, has discovered the importance of professional development that is focused squarely on increasing teachers' content and pedagogical content knowledge and teaching skills. She has also observed a shift by decision-makers to seeing and believing that the purpose of professional development is to enhance learning of challenging content for all students. Murphy feels

How do we know if our staff training program did, in fact, bring about our intended learning outcomes?

that this shift has led to increased accountability and responsibility of professional development programs to better equip teachers to teach a rigorous curriculum to all students and to ensure that students have every opportunity to meet the highest standards. So, her work clearly establishes professional development as a tool to focus on building the knowledge and skills of teachers which becomes a link to enhanced student outcomes. Murphy also believes that the schools that carefully think about what kinds of professional development programs they need, based on student learning needs, are most successful in improving student learning. These schools analyze data to find out where their students are not succeeding and develop professional development plans related to enhancing teaching and learning in those areas. Schools can measure their return on investment of professional development dollars by gathering evidence of change in teacher practice and student learning outcomes.

This incremental effect professional development has on improving individual teacher skills designed to improve individual student learning can have a multiplying effect. Chapman and Harris (Winter 2004) believe that professional development will go beyond the classroom and impact total school improvement. They feel that the concept of the "teachers as learner" is key to school improvement and school effectiveness. They found that professional development was one of the most important factors in securing school improvement. And the opposite was as also true; a lack of investing in staff development over time resulted in an erosion of professional confidence and capacity and a major barrier to improving schools.

So, it would seem that there is compelling research to show that some kinds of professional development are effective in promoting student achievement.

How Can We Judge the Effectiveness of Our Professional Development?

How do we know if our staff training program did, in fact, bring about our intended learning outcomes? Let's examine a typical, hypothetical case. In this short vignette, the school in our example has had, for the past several years, fourth grade math scores much lower than scores in similar schools in the area. The summer before our fictional school opened this year the fourth grade teachers received intensive training to help them understand how to align their lesson planning and classroom assessments to the new state content standards for fourth grade mathematics. And, to no one's complete surprise, fourth grade math scores slowly, but steadily improved over the course of that following school year. But did that training, in fact, produce those results? What else might have caused the improvement? Was the district involved in other, unrelated initiatives that should be considered? Did this year's fourth grade class have a different third grade curriculum, or teacher, which might have contributed to their progress? Where some influences outside the school at work, such as a change in demographics? We all know the multiple variables, both in and outside the school, which can affect learning positively and negatively. To isolate one factor as a cause for an observed effect becomes difficult, at best, for the typical school leader and decision-makers.

Beth Kubitskey (2003) and her colleagues agree that it is problematic to tease out learning outcomes with direct absolute correlation to teacher action. When the student experiences a curriculum designed to offer multiple methods of learning, isolating one activity and trying to link the learning to that activity is difficult. In addition, the teachers are offered multiple forms of professional development which complicates linking teacher learning to any one of the training experiences. However, these researchers feel that, although weakly linked, if we identify a positive learning outcome that may have come from a given activity, we can use this information to suggest continual inclusion of said activity in professional development. Likewise, students' failure to meet the learning objective may indicate a need for modifications in the professional development linked to the teacher's implementation of the activity being studied. In other words, if you start to see learning gains among the targeted student population, keep doing what you were doing.

Guskey (March 2002) has found that effective evaluations of professional development require the collection and analysis of five critical levels of information:

- *Participants' Reactions* - Did the participants value the training? This is the lowest evaluation level.
- *Participants' Learning* - What did the participants learn as a result of the training? This is a bit more important, but it doesn't tell us what happened in the classroom.
- *Organization Support and Change* - Did the school or district leadership advocate for the training and support the participants in the training and in their application of these new skills in the classroom?
- *Participants' Use of New Knowledge and Skills* - Did the participants actually apply their new skills and knowledge in the classroom?
- *Student Learning Outcomes* - Most importantly, did the students improve as a result of their teachers' professional development (46-49)?

"... culture enhances professional learning when teachers believe professional development is important, valued, and "the way we do things around here."

This hierarchy of observed effects, from simple participant perceptions to the more complex actual student learning gains, is a convenient way to think about what kinds of evidence we must gather. Guskey explains that there are three important implications that stem from this model for evaluating professional development

- Considering each of these five levels is important
- Tracking effectiveness at one level tells you noting about the impact at the next
- Planning professional development to improve student learning requires that the order of these levels must be reversed (50).

For the last implication, we must consider that to have truly effective professional development designed specifically to improve student learning, we must first decide what is it that we want are students to learn and how will we know if they learned it? Guskey believes that if you don't know where you are going, it's very difficult to tell whether you've arrived. But if you clarify your goals up front, most evaluation issues fall into place.

> Several years ago Stephen Covey (1989) said that it is critical to begin with the end in mind (which) means to start with a clear understanding of your destination. It means to know where you're going so that you better understand where you are now and so that the steps you take are always in the right direction... How different our lives are when we really know what is deeply important to us, and, keeping that picture in mind, we manage ourselves each day to be and to

> do what really matters most. If the ladder is not leaning against the right wall, every step we take just gets us to the wrong place faster. We may be very busy, we may be very efficient, but we will also be truly effective only when we begin with the end in mind (98).

Another aspect of professional development makes it difficult to assess the effectiveness of our training. In a vast number of instances, teachers who attend professional development training sessions do so under their own initiative. In other words, they are not compelled to attend by either district or school leadership. While many teachers do participate as a result of district mandate, William Bobrowsky and his associates (March 2001) found that most research on professional development effectiveness is based on studies where the participants were self selected for training. Bobrowsky calls these willing participants "volunteers" who tend to be innovators, actively seeking new ideas that challenge their present thoughts on teaching and learning. They do not see change as a flaw in themselves and are risk-takers. The teachers who do not participate, he calls "non-volunteers". These may not be unwilling, but are likely unimpressed by the types of professional development currently offered them by research groups. Bobrowsky believes that professional development needs to resonate with the beliefs and attitudes of more than just the volunteer populations of teachers. Professional development needs to reach out to those teachers who have differing beliefs. He is further concerned about the professional development that is created and delivered by the individual school districts. His review of the research shows that this type of training tends to be piecemeal, fragmented and is not individualized to fit the needs of teachers.

Professional development centered on student achievement goals is meaningful to teachers, enabling them to base their instructional decisions on solid evidence of what students need.

If we know that certain types of professional development increase the chances that students will actually perform better and if we know how to begin thinking about evaluating our own professional development programs, what are the shared characteristics of these effective training programs?

Professional Development Characteristics Linked to Student Improvement

What can educators and decision-makers do to become better consumers of professional development? And, what are the characteristics of professional development offerings that are most closely correlated to improved student learning and what must schools do to help increase the chances that this training will, in fact, lead to improved student achievement? We know that a teacher satisfaction rating is the weakest indicator for predicting student outcomes, but there are other things that we can and should consider. For instance, Barry Fishman and his associates (April 29, 2000) have found that the school culture that plays a vital role in enhancing the effectiveness of any faculty training program. Fishman and his colleagues contend that culture and norms of the system or school are related to the settings in which the innovation is to be carried out. For instance, does the school principal support the curriculum? These factors will influence the attitudes of teachers toward the innovation. If the principal attends the summer work session, a teacher is likely to think, "This is important to him/her, I should find a way to make this work!" (4). As a result, the school leadership plays an important role in setting the tone and making training an important part of school life.

Kent Peterson (Summer 2002) affirms the value of school culture in establishing an environment conducive to effective professional development. Peterson reports that school culture enhances or hinders professional learning. Culture enhances professional learning when teachers believe professional development is important, valued, and "the way we do things around here." Staff learning is reinforced when sharing ideas, working collaboratively to learn, and using newly learned skills are recognized symbolically and orally in faculty meetings and other school ceremonies. The most positive cultures value staff members who help lead their own development, create well-defined improvement plans, organize student groups, and learning in a variety of ways. In some schools, professional development is not valued, teachers do not believe they have anything new to learn, or they believe the only source for new ideas is in trial-and-error in one's own classroom. And Lynda Abbott's research (June 2005) supports the fact that the social context of a school is important for change. Teachers in her study were generally acting as agents for change to other teachers and being actively involved in conducting formal or informal professional-practice instructional sessions for others.

Principals and other school leaders can and should shape school culture. As Hessel and Holloway (2002) point out,

> The school leader models the very essence of behaviors expected from teachers and students. The leader's knowledge base, enthusiasm, skill, and modeling inspire others to achieve at high levels. The leader encourages and models life-long learning and striving for excellence. Nothing less than the best effort is what is expected every day in every way from every person... The culture created through this (interdependent, connected) relationship fosters the achievement of the school's ultimate mission, success for all students (51).

Once a culture is established in our school that values both student and teacher learning we can address the specific elements of our professional development program that will impact student achievement. This type of training will focus on helping teachers change to address school and student goals in an evidence-centered way. This means we will want to engage teachers in training sessions that will be based on both teaching and student standards. We want teachers to begin to think about such things as: "How will my performance in the classroom change as a result of this training?" "What standards should I be addressing, and how will I know if I attained those standards?" We also want the teachers to think about how their students will be different as a result of this training and will they be able to provide evidence of those changes and make judgments about that evidence. As Stephanie Hirsh (Winter 2004) points out, planning professional development to address how individuals must change to achieve district's goals is most effective when the training is results-driven, standards-based, and focused on educators' daily work. In a results-based focus, however, the training outcome is measured against what gains we expect in teacher or student performance. And finally, instead of considering the one shot workshop, one time a year professional development that focuses on the daily work of educators assures that the training is job embedded and relevant to practice, aligned with district or school goals, and uses actual student achievement data to inform the training and future planning.

Sparks and Hirsh (1997) believe that a fundamental shift must occur in the way most districts think about professional development. For instance, they advocate that professional development must:

- Be driven by clear, coherent strategic plans
- Focus on student needs and learning outcomes
- Include multiple forms of job-embedded learning
- Provide opportunities for study by teachers of the teaching and learning processes
- Include continuous improvement in performance for everyone who affects student learning
- Consider professional development as an indispensable process without which schools cannot hope to prepare young people for citizenship and productive employment (12 -16).

Holloway's review of the literature (November 2003) establishes the importance of using (student) performance data because that allows educators to focus their valuable and limited professional development resources on the specific learning needs of students. Professional development centered on student achievement goals is meaningful to teachers, enabling them to base their instructional decisions on solid evidence of what students need. More important, such professional development supports the goal of ensuring the success of all students.

In some schools, professional development is not valued, teachers do not believe they have anything new to learn, or they believe the only source for new ideas is in trial-and-error in one's own classroom.

Using student performance data becomes effective, according to the research of Lawrence Ingvarson and his colleagues (January 29, 2005) when, training sessions provide opportunities for teachers to focus on what students are to learn and how to deal with the problems students may have in learning that subject matter. In addition to paying attention to student data, they found that other factors proved valuable in effective professional development. For instance, effective training programs:

- focused on research-based knowledge about student learning of content
- included opportunities for teachers to examine student work collaboratively
- provided time for teachers to actively reflect on their practice and compare it with high standards for professional practice
- engaged teachers in identifying what they needed to learn, and in planning the learning experiences that would help them meet those needs
- provided time for teachers to test new teaching methods and to receive follow-up support and coaching in their classrooms
- included activities that led teachers to deprivatize their practice and gain feedback about their teaching from colleagues (15-16).

One of the key ingredients that the Ingvarson study found in effective professional development was follow-up and coaching after the training as the teacher attempted to apply this new knowledge to practice. Although Ingvarson encourages feedback after training, his study found how rarely designers built in opportunities for feedback and coaching in the workplace, despite research on their centrality to learning new and complex skills. Other key ingredients are time to think, analyze and talk about the specifics of what is going on in classrooms and what students are doing and learning. As other researchers have reported, Ingvarson also advocates a positive school learning culture were effective school administrators expect evidence of professional development an act in ways that demonstrate they value teacher learning.

The right kinds of professional development for both teachers and school leaders can directly contribute to improved student performance.

Conclusion

The right kinds of professional development for both teachers and school leaders can directly contribute to improved student performance. And, schools can become advocates for investment in increased professional development by capturing evidence of its effectiveness to show a return in learning of their investment in time and money. Research is beginning to emerge that points to some key ingredients in professional development for improved student learning. Some of these contributing factors include:

- focusing on teacher content knowledge and teaching skills
- considering student learning goals and the training outcomes in an evidence-based way
- creating a supportive culture for a learning community among all members of the community
- using student data to inform professional development planning and as part of the training itself
- embedding the training in the daily work of the teacher
- sustaining the training over time
- allowing for feedback and coaching
- providing opportunities for the teachers to participate in the planning of their training and to reflect on their practice

Effective, research-based professional development can contribute to the continuum of teacher training and support from preservice training to advanced practice. More importantly, however, high quality professional development can translate into improved student achievement and overall school effectiveness.

References

Abbott, L. (June 2005). The nature of authentic professional development during curriculum-based telecomputing. *Journal of Research on Technology in Education*, v7, p379

Bobrowsky, W., R. Marx. And B. Fishman. (March 2001). The empirical base for professional development is science education: Moving beyond volunteers. Paper presented at NARST 2001, St. Louis, MO, March 26, 2001. Available online at http://www-personal.umich.edu/~fishman/papers/Bobrowsky-NARST2001.pdf

Chapman, C. and A. Harris. (Winter 2004). Improving schools in difficult and challenging contexts: Strategies for improvement. *Educational Research,* 46(3), 219-228.

Covey, S. (1989). *The 7 habits of highly effective people*. New York, NY: Firestone.

Fishman, B., S. Best, J. Foster, and R. Marx. Fostering teacher learning in systemic reform: A design proposal for developing professional development. Paper presented at NARST 2000, New Orleans, LA, April 29, 2000. Available online at /http://www-personal.umich.edu/~fishman/papers/Fishman-NARST2000.pdf

Guskey, T. (March 2002). Does it make a difference? Evaluating professional development. *Educational Leadership*. Alexandria, VA: Association of Supervision and Curriculum Development. 59(6), 45-51.

Guskey, T. (Winter 2005). Five key concepts kick off the process. *Journal of the National Staff Development Council*. Oxford, OH: NSDC 26(1), 36-40.

Hessel, K. and J. Holloway. (2002). *A framework for school leaders: Linking the ISLLC Standards to practice*. Princeton, NJ: Educational Testing Service

Hirsh, S. (Winter 2004). Putting comprehensive staff development on target. *Journal of the National Staff Development Council*. Oxford, OH: NSDC 25(1), 12-15.

Holloway, J. (November 2003). Linking professional development to student learning. *Educational Leadership*. 61(3), 85-87.

Ingvarson, L., M. Meiers, and A. Beavis. (January 29, 2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives*. 13(10). Available online at http://epaa.asu.edu/epaa/v13n10/

Kubitskey, B., B. Fishman, and R. Marx. (2003). The relationship between professional development and student learning: Exploring the link through design research. Paper presented at annual meeting of AERA 2003. Available online at http://www-personal.umich.edu/~fishman/papers/Kubitskey_AERA2003.pdf

Murphy, S. (Spring 2005). Changing perspectives in professional development. *Science Educator*. 14(1), 9-15.

Peterson, K. (Summer 2002). Positive or negative. *Journal of the National Staff Development Council*. Oxford, OH: NSDC 23(3), 10-15.

Sparks, D. and S. Hirsh. (1997). *A new vision for staff development*. Alexandria, VA: Association of Supervision and Curriculum Development and Oxford, OH: National Staff Development Council.

John H. Holloway is school leadership director, Professional Development Group, Educational Testing Services, Rosedale Road, M/S 18D, Princeton, NJ 08541. Correspondence concerning this article may be sent to jholloway@ets.org

Adrienne Gibson and John Wallace

Teaching and Assessing Science For Understanding: Managing the Accountability Dilemma

The authors report on findings from a year-long study of how teachers determine student understanding and why they employ the instructional and assessment strategies that they do.

Introduction

In the United States, science teachers are challenged by the National Science Education Standards (NSES) to "select teaching and assessment strategies that support the development of student understanding and nurture a community of science learners" (National Research Council, 1996, p. 30). Such pedagogies, broadly based on the notions of constructivism and teaching for understanding, encourage teachers to be flexible and to shift the pace and focus of instruction on a daily basis to meet the needs of students. These reforms acknowledge that students construct their own understandings of any given content focus. Assessment practices in a classroom oriented towards understanding must provide ongoing information that allows a teacher to be responsive to individual needs (Perrone, 1998). In a classroom where instruction has been differentiated to meet the needs of individual students, formative or ongoing assessment becomes a critical factor in understanding how to adjust instruction as students begin to achieve the desired understandings. In such a classroom assessment becomes "ongoing and diagnostic … today's means of understanding how to modify tomorrow's instruction" (Tomlinson, 1999, p. 10)

Paralleling the NSES reforms, there is also a strong pressure on teachers to prepare students for national and state standardized tests. Hardy (2004) indicates that there is strong public support for the current political pressures on schools to ensure their students will achieve high test scores. The general public response to the vision defined by the No Child Left Behind Act has been favourable (NCLB, 2001). The No Child Left Behind law now requires that by the year 2007 states determine student achievement in science at least once in each of three grade ranges (NCLB, 2001). Scores on these tests are seen as general indicators of student progress and comparisons of national, state, district and, sometimes, individual school scores are used to evaluate the quality of schools and the effectiveness of specific programs and teachers.

Inevitably, the effects of these national testing regimes are being felt at the classroom level where teachers feel compelled to 'teach to the tests'. However, such teaching is built on different premises about student learning, pedagogy and assessment than teaching for understanding. The emphasis when teaching to the tests is likely to be towards knowledge acquisition rather than knowledge construction, direct rather than flexible pedagogies and summative rather than formative assessment practices.

Analysis of student learning should serve as the basis of instructional planning.

Herein lies a fundamental dilemma for teachers as they try to accommodate these countervailing pressures, to teach and assess for understanding *and* improve student performance on standardized tests. Teachers are being asked to be accountable in two ways: to students (and their understandings) and to the public and to school boards (for improving overall student test scores). We call this the accountability dilemma. Wilder and Shuttleworth (2005) also refer to the dilemma that teachers address daily as they attempt to meet the vision of the NSES by providing opportunities for inquiry based learning and also meet the demand of state and federal testing programs for content coverage. In this study we examine how five teachers from one school district managed to teach for understanding in the face of this dilemma. We report on the findings of a year-long research project investigating how these teachers determined students' understanding and why they employ the instructional and assessment strategies that they do. Our research question is: How and why are teachers teaching and assessing for understanding within the context of the accountability dilemma?

Background

Constructivist learning theory asserts that learning is an active process of constructing meaning and that knowledge cannot be transferred as a complete assemblage from one individual to another. From a constructivist perspective, students are understood to use prior knowledge and understandings to build new knowledge and understandings. Constructivism, as a theory of how learning is achieved, does not delineate a blueprint for action on the part of teachers. However science teachers, using a constructivist referent, are called upon to provide opportunities that challenge students' existing knowledge and help them develop new understandings of the world around them. Much has been written about strategies that teachers might use if they are teaching for understanding, often using constructivism as a referent (Brooks & Brooks, 1999a, 1999b; Llewellyn, 2005; Prawat, 1989; Tobin & Tippins, 1993; Windschitl, 1999, 2000; Yager, 1996). What is clear is that, if instruction is designed using the tenets of constructivism, there must also be a concurrent shift in strategies for assessing student understanding (Holmes & Leitzel, 1993).

Many science reform initiatives have centred on the need to teach science for the understanding of concepts rather than for simple recall of facts without meaning. A conceptually-oriented view of teaching includes the attributes of focus and coherence, negotiation, and analysis and diagnosis (Prawat, 1989). It is this third attribute that science teachers, in particular, would recognize as being critical to the progress of their students. Analysis of student learning should serve as the basis of instructional planning. Ongoing assessment of student understanding by the teacher provides information about the obstacles students are encountering while negotiating their way toward conceptual understanding. Wiggins and McTighe (2005), for example, encourage teachers to adopt a process of backwards design in creating curricula materials focused on teaching for understanding. They challenge teachers to "gather lots of evidence along the way" and explain that "the assessment of understanding should be thought of in terms of a collection of evidence over time instead of an event" (p. 152).

Constructivism, as a set of beliefs about knowing and knowledge, can be used as a referent for teachers to determine the learning potential of a given classroom situation. A constructivist view of learning has consequences not only for how one views an understanding of science content but also for how the role of a teacher is perceived. According to Duit and Treagust (1995, p. 52), under this orientation, "the teacher is viewed as a facilitator of knowledge and not as a person who transfers knowledge to the brains of students." Teachers need to decide what experiences should be provided in order to facilitate learning and how the learner can represent what is already known to give meaning to the new experiences. Brooks and Brooks (1999a) identify five principles that teachers who are interested in creating constructivist classrooms need to understand: "teachers seek and value their students' points of view, classroom activities challenge students suppositions, teachers pose problems of emerging relevance, teachers build lessons around primary concepts and big ideas, teachers assess student learning in the context of daily teaching" (p. IX-X).

An important guiding framework for science education in the United States is the National Science Education Standards (NSES), published by the National Research Council (NRC, 1996). This document incorporates standards for science teaching, professional development of teachers, assessment in science education, science content, science education programs, and science education systems. The NSES take the position that if teachers are to be prepared to teach for understanding, they need strong content and content specific pedagogical knowledge. And, importantly, for the purposes of this study, they require a clear understanding of how students learn. While the NSES do not specifically espouse any particular model for instruction, many of the teaching standards proposed in the NSES use language also found in the literature on teaching for understanding and constructivism.

Although the NSES support a variety of performance-based assessments, a countervailing trend in both the political and educational arenas is towards a greater use of standardized test scores as a measure of student achievement. The most common purposes of state testing are accountability for the schools and school districts, instructional improvement, and program evaluation. A less common purpose is student diagnosis and placement. Widespread testing for accountability purposes has increased following the 2002 enactment of the federal *No Child Left Behind* legislation. Throughout the United States, there are state tests for students in mathematics and language arts, tests for science and social studies achievement are being developed and many states now have high school exit exams. In 1995, Ryan and Miyasaka reported that multiple-choice tests were the most common format (70%) for state tests, with other formats such as performance assessment (28%) and portfolios (18%) less evident. By contrast, within the classroom, nonmultiple-choice testing formats were predominant (Barton & Coley, 1994).

We have to make connections between ideas if greater meanings are to happen.

In summary, teachers are working within a complex and contradictory teaching and assessment milieu. The widely lauded National Science Education Standards (NSC, 1996), drawing on research into teaching and learning science, advocate a range of classroom, performance-based, normative as well as summative, measures to assess student understandings. The popular standardized testing movement draws from a different set of theoretical assumptions about teaching and learning and school reform. This movement has resulted in the widespread administration of external, summative, uniform multiple-choice type tests. In their classrooms, teachers are responding to both pressures. While employing mainly nonmultiple-choice techniques to monitor understanding, they are also preparing students to sit for multiple-choice state tests. This study of classroom assessment practice examines how teachers manage to inhabit these two assessment worlds, and how they balance the dilemma of having to account to different groups in different ways.

Research Design and Procedures

This data for this article were collected by the first author as part of a year-long qualitative study of five science teachers from the Sonora Foothills School District—Ms. Springer, a multi-age teacher of grades one through three, two fifth grade teachers, Ms. Frederick and Mr. Gorman, a sixth grade teacher, Ms. Richardson, and a tenth grade teacher, Ms. Cornette (district and teacher names are pseudonyms). The Sonora Foothills District had a declared policy of implementing the NSES as well as improving student scores on state and national standardized tests. Multiple methods of data collection included observational field notes, teacher journaling, and semi-structured interviews. During the course of the study, Adrienne worked alongside the teachers as a participant observer (Adler & Adler, 1994). The data were reported in the form of individual case studies, incorporating lesson vignettes and teacher reflective commentaries. The case studies were analysed to identify common themes about teaching for understanding.

Yin (1994) defines a case study as "an empirical inquiry that investigates a contemporary phenomenon within its real life context" (p. 13). In this study, the contemporary phenomenon was teacher perceptions of student understanding within the real-life context of the science classroom and the political and profession pressures that teachers were experiencing. The study fell into the category of a collective case study as several cases were studied jointly in order to inquire into the phenomena of teacher perceptions. Individual cases were chosen with the belief that understanding them might lead to a better understanding of a larger collection of cases (Stake, 1994). We see the analysis of these cases as a small step toward understanding on a general level the strategies teachers use to assess student understanding and their success in doing so.

In the interest of brevity, in this article we present a full account of one of the cases, that of Ms. Joan Springer. While we do not claim that the case is representative of all five cases, it does capture most of the issues which emerged during this year-long study. The case includes an introduction and narrative vignette (Polkinghorne, 1995) of one of Joan Springer's lessons, narrated by Adrienne, followed by a reflective commentary on the central issue of teaching for understanding, narrated by Joan. The case is followed by a discussion of the issues emanating from this and the other four cases, supported by excerpts from the other cases where appropriate.

The Case of Joan Springer

Ms. Joan Springer was an experienced elementary school teacher, having taught in kindergarten - sixth grade classes for the last 28 years. She was also a teacher leader in the

Sonora Foothills District in the area of staff development and, for several years, helped teach a requisite class for all new teachers that covered Madeline Hunter's Essential Elements of Instruction (Gentile, 1988; Hunter, 1979). During the five years previous to this study, Ms. Springer taught a multi-age, first through third grade, class.

Within her multi-age classroom, Ms. Springer worked to build a sense of community among her students. When in her classroom it was very common for me to observe older students helping younger ones without prompting; sharing their knowledge and understanding of various concepts and helping the younger ones learn basic skills. By adhering to the multi-age concept that students learn best by learning from many different people in many different ways and with the help of many classroom visitors and volunteers, Ms. Springer managed to balance the range of needs presented by her multi-age students.

The following case study story focuses on one of Ms. Springer's lessons in a unit on electricity. Although this was actually a unit in the third grade curriculum, all the students in the room participated in the lesson. I observed the case study story lesson close to the beginning of the electricity unit. During most of the lesson I sat with the students and scripted as much of the conversation between students and teacher as possible. When the students were assigned to work in their learning groups my role changed from observer to participant and I was no longer able to record the students' comments. Immediately after the lesson, I left the classroom and wrote as many of the students' words as I could remember in my own journal. I also recorded my impressions and questions about the lesson so that I would later be able to discuss them with Ms. Springer. Within a day or so after my observation she sent me her journal notes, including her thoughts about what she perceived to be a dilemma posed by the electricity unit as a whole. Later, when we had an opportunity to talk about it, Ms. Springer related the difficulties she was experiencing as she tried to maintain the integrity of her multi-age program while, at the same time, planning for appropriate science instruction for all of the students in her classroom.

Teaching Vignette: "Atoms Get Recycled"

Students entered Ms. Springer's class as soon as the bell rang. Class began with the students reciting the "Pledge of Allegiance" and then singing the class song. Ms. Springer told her students that she would be reading them a story about electricity before they started on their 'theme' (science) project for the day. She told them that the story would be about electrons traveling around making light and heat. The electrons would also be traveling thorough wires during the story.

The students sat on a rug and listened quietly as Ms. Springer read. Immediately after the story was finished, they raised their hands to make comments. A third grader said, "I'm thinking about yesterday and I think atoms get recycled."

Mrs. Springer responded, "Atoms get recycled? What a wonderful concept and good thinking. What did the electrons in the story do?"

Another student answered, "Electrons go differently. Only one atom jumps. Then it dies and another one jumps."

"What else did you learn from the story?" Ms. Springer asked the class.

Again a third grader answered. "I learned that at the power plant they send out a whole lot of electricity and when it gets to our houses it lights the lights."

"That fits into what I want to talk about today," Ms. Springer replied. "We are going to add sources of electricity to our web." Ms. Springer frequently used a concept map or web design to illustrate how various parts of a lesson interrelated. At this time she pointed to the web that had been started the day before. "Sources are where electricity comes from. There are three places. One is the sun. Do you remember the others from yesterday?"

"Water," answered one student.

"Wind," another said.

"What did I ask you to write about yesterday? What you think about electricity?"

This time a second grader began to get in on the discussion. "Could we read what we wrote?"

Another second grader joined in with, "I learned that when electrons jump we get electricity." Several students shared the comments that they had written in their journals on the previous day after seeing a Bill Nye movie about atoms and electricity. Many repeated the idea that electrons jump from atom to atom.

"Do you remember what I said we were going to do next?" Ms. Springer then asked. "Each person is going to do an individual report and each group is going to do a project as a group. Your reports can be about any part of our electricity web that you are interested in. The web includes the words atoms, sources, how it works, and safety centered on the word electricity. In order to write about one of these four parts we need to have a little more information about atoms. Can you understand about electricity if you don't know about atoms?"

"No", a third grader replied.

Ms. Springer continued with a brief discussion of the structure of atoms. "So we have to learn about atoms. Look at the back of your hand. Do you see little sections? These are like cells and they are made of molecules. Molecules are made of atoms. They are very, very tiny. You can't see them with just your eye. Do you remember what an atom looks like?"

A third grader answered quickly, "Atoms have a nucleus inside and that is made of protons and neutrons and electrons around the outside."

Another question from the teacher, "And if electrons jump from atom to atom in a circle that is closed what do we have?"

A third grade student answered, "A circuit, electricity."

Ms. Springer then explained to her students that atoms are invisible to us and students would have to take the scientist's word for the fact that they exist. "This is one of the things that has to be taken on faith." She then showed the students a diagram of an atom. "It is just like you have described. Look how much space there is between the nucleus of an atom and the electrons. Atoms make up your body and your body is 99% space."

"Take a minute before we go on and write about an atom in your journal. You little ones can draw and then put labels on the parts. The second and third graders need to write some sentences. I'd also like you to write about an "aha" moment that you had from the video, something that you now understand because you watched it. What were some of your "aha" moments?"

The students' responses included "energy takes lots of power," "a turbine makes water to electricity" and "I want to find out about kinetic energy so I understand it."

"How long do you think it will take to understand all there is to know about electricity?" asked the teacher.

A first grade student who had been silent during rest of the conversation didn't wait to be called on. "All your life!"

After a few minutes had passed and all the students had added something to their journals Ms. Springer continued with the lesson by giving directions for the remainder of the class period. "OK, when you're done with your journals you are to get in your learning groups and start working on your reports. You can also talk about your project with your group if you wish." "Remember that you older students are responsible for your own reports as well as working with the younger kids when they need help." Students were also instructed to decide on a group project with each student to be given a task appropriate to his or her grade level abilities and interests. Each group was composed of four students including at least one third grade, one second grade and one first grade student.

To me, understanding is when students demonstrate they know the concept or skill by correctly performing it in a similar but not exactly the same way.

At this point in the lesson I stopped scripting the dialogue and began to help various children find books that they wanted to use. A third grade girl asked me for help in locating information about how a hydroelectric dam worked. Together we found a book with a good picture and accompanying diagram with an explanation. I let her explain the process to me until she became frustrated with trying to understand how turning the turbines created energy that was stored in what looked like a big battery in the diagram. I reminded her of what her classmate had said earlier about it possibly taking a lifetime to understand everything about electricity and suggested that she did not need to understand every detail in third grade in order to do her report. She was visibly relieved and confided that she had been to a hydroelectric dam with her family and after this report she would be able to explain to them just how it worked.

As I continued to circulate around the room I had other opportunities to ask the students about what they had learned or were in the process of learning. One group, whose project was going to be about electrical storms, shared the following insights with me. A third grade student related that "electricity comes from storms and atoms, electrons jump from atom to atom making electricity"; another third grade student shared that "atoms are electricity, they are so small that we can't see them but scientists can and they know they are there"; the second grade member of the group explained "atoms are electricity, they jump through a wire to the light bulb. The one first grade student in the group had been following the direction of the older students during their project activities. He had spent most of the time carefully drawing a picture of an atom including a nucleus with electrons in a ring around the outside. When it was his turn to share his learning he smiled at me and said, "I love atoms. I'm doing my report on them. They are very small. I have to use my imagination to see them."

The lesson illustrated by the case study story raised many questions in my mind about how Ms. Springer made assessments of what the students had actually learned during the course of the lesson. She and I discussed these questions and others by an exchange of journal entries and in follow-up conversations. In addition, during the field study, I was able twice to question Ms. Springer extensively about her background in teaching, the strategies that she finds most useful in teaching science and her perceptions of what her students understood during the lessons that I had observed. The following commentary was constructed almost entirely from Ms. Springer's words during our conversations and interviews. It represents her own reflections on her teaching and on the methods she relies on to determine what her students are accomplishing.

Joan's Reflective Commentary

"I can't," is rarely heard in my classroom. I have tried to create a classroom culture in which all the children feel like they can do whatever they are asked to try. I'm finding that the longer I teach multi-age the more that is happening. We are a group for a long time and it is important to create a sub-culture of sharing that encourages each student to support the others as they learn. One of the things we have made a point of is that our information or knowledge should not be kept a secret. It should be shared so we can all build on it.

When I am introducing a new science topic or unit I like to use a "web" type diagram, much like a concept map. It traditionally includes the parts of the topic we are going to explore and what I want the students to find out. I use the same idea of a web to help the students build bridges between the different units that we work on. My students build mental bridges not only from unit to unit but also from year to year. The kids will make statements to each other like "... remember last year, how we did that? Well this is how it fits."

Another thing that I think is useful for teaching is a clear understanding of Bloom's Taxonomy of critical thinking skills. The ability to use that taxonomy to ask questions and elicit responses is critical for discussion within my multi-age group. I expect responses at a higher level of thinking from my older students than from the younger ones. I try to deliberately phrase questions so that each child will have to stretch his or her thinking.

I think it is vitally important that I am aware of my students' existing beliefs and understandings. We have to make connections between ideas if greater meanings are to happen. If students don't have a base to connect with, learning doesn't take place. As a teacher I have to know if everyone is on the same track or someone will get lost along the way while I am going on and on. You need to help students get back on the right track as soon as you can so they don't get lost. You have to talk with kids all the time to see what they understand. I try to check for understanding constantly by talking to my students, watching them work and asking questions. Sometimes I am very methodical about checking at specific stages in a long-term project. Typically my questions might be "What is ...?; How do we know that ...?; Where do we go from here ...?" With my multi-age kids a younger one might not know how to answer each question but will benefit from having the correct answer repeated and explained by an older child.

I am not a proponent of standardized tests as a means of determining how well my students are doing. I don't think the format provides a fair test of what children can do. To me, understanding is when students demonstrate they know the concept or skill by correctly performing it in a similar but not exactly the same way. This could be through conversation or demonstration. They are successful when they can bring the learning back and apply it to a different yet similar situation or build onto that learning to go to the next step. I think that levels of understanding are different. What one person understands at an early stage should deepen and expand with more experiences with the same or similar concept.

The most effective way for me to assess the progress of my students is to listen to them on an individual basis. Final assessments come from presentations and projects. Even when there are group projects, each student is expected to complete an individual component on his/her level. Overall, I think my students understand a great deal about their lessons. People ask about multi-age all the time. They don't understand how you can teach a first grader and a third grader at the same time. I wish they could see the students' presentations. First graders might stand up and say a few words about plants while third graders explain photosynthesis. This year, during presentations to parents, one of my students announced that he didn't wish to share his report. Instead, he talked about how everything is connected. He clearly was able to build bridges between topics he had discussed in class. This was just an ordinary kid that loved what we had been doing and was involved in his own learning.

Discussion

The purpose of the study was to describe how and why teachers are teaching and assessing for understanding within the context of the accountability dilemma. The data, exemplified in the above case of Ms. Springer, included observations of the teachers' instructional and assessment strategies. In analyzing the data, we identified commonalities of thought and action across the five case studies. We found that it was not possible to separate teachers' practices into two disconnected sets of actions, one being of instruction and the other of assessment. In each case study, assessment of student understanding was embedded in instruction, as was instruction, in some instances, embedded in assessment. As we analyzed the data we were able to identify themes that were both specific to each individual teacher's practice and common among all of the participating teachers. The themes included the link between teaching and assessing for understanding, the tension between professional and political assessment expectations, the preference for informal information on student achievement, and teachers' capacity to manage a dilemma of pedagogical choices.

The link between teaching and assessing for understanding

The outcomes of this research confirm that each of the teachers in the case studies developed their own perceptions of their students' understanding by using a number of different assessment strategies. These strategies were embedded in their individual practices of instruction as well as assessment. As a result it was difficult to separate their instructional strategies from their assessment strategies.

The study teachers were observed using a wide range of techniques, including journaling activities, questioning strategies, group projects, one-on-one interviews, concept maps, activity reports, and student presentations to determine their students' levels of understanding during and at the end of the instructional units. As exemplified in the "Atoms Get Recycled" vignette, teachers frequently used information gained from these sources to adjust their instruction in order to meet their students' needs. Prawat (1989) identified teaching for understanding by its "highly analytic and diagnostic nature" (p. 324). Just as Ms. Springer encouraged her students to explore their ideas about atoms, the other study teachers favored a facilitation role over a transmissionist role. In the process of facilitation they used the analysis of student understanding for instructional decision making such as the addition of alternative activities to aid the progress of students with different learning styles. Ms. Springer employed this strategy when she added activities to her electricity unit because some of the students were having trouble understanding the concepts she wanted them to learn. Another of the study teachers, Mr. Gorman, employed a similar strategy when teaching a lesson on Bernoulli's principle.

Final assessments employed by the five study teachers included performance assessments, project presentations, one-on-one interviews and student-led parent conferences as well as teacher-generated paper and pencil tests. For example, Ms. Springer had her students complete independent projects on electricity and present their work at a parent gathering. During these processes the teachers often provided instructional feedback comments. Ms. Springer did this while the students were working on their projects. Another of the teachers, Ms. Fredericks, used one-on-one final assessment interviews to teach students the concepts they were struggling with. In these instances the focus of assessment was on student understanding of concepts and the processes of reasoning rather than on the production of 'correct' standardized answers.

The tension between professional and political assessment expectations

The teachers in the Sonora Foothills School District were challenged by the district goal to be "the number one school district in the nation … as evidenced by test scores." At the time of this study, teachers were offered incentive pay bonuses if they were able to meet specific short-term goals in order to make progress towards the district goal. The science teachers who participated in this study often talked about the difficulties in reconciling their preferred instructional and assessment practices and the testing parameters established by the school district.

Teachers felt that they were subjected to contradictory direction about how to determine the level of understanding that students were achieving in their classrooms. In each case they felt they were receiving conflicting signals from the district administration in terms of how to adequately evaluate the progress of their students. Teachers were encouraged by the district to participate in workshops and training including several days of training in Madeline Hunter's elements of instruction, multiple

Teachers' knowledge of science content and science content pedagogy as well as their personal knowledge and experience impacted their instructional and assessment decisions.

intelligences, brain-based learning, multi-age classroom instruction, and the use of performance assessments. Ms. Springer, for example, took a lead role in this program. The district curriculum had been rewritten shortly before this research began. The teachers involved in the curriculum revision used the NSES as the framework for the revision process. One of the study teachers, Ms. Cornette, a high school chemistry teacher worked with colleagues who had had extensive training in using learning cycles to teach for understanding. Many of those opportunities supported practices recommended by advocates of constructivist teaching. At the same time, however, the teachers were being encouraged to develop teaching strategies deliberately focused on raising the standardized test scores.

Within their classrooms the teachers planned for science instruction and assessment in light of these contradictory directions. They fulfilled their responsibilities as they understood them to be in terms of administering standardized tests but most of the teachers did not seem to place much value in the information that might be gained from the test results. In her commentary Ms. Springer dismissed the validity of the mandated standardized tests in determining student understanding and achievement. Another teacher, Mr. Gorman, was also concerned that scores on "the district tests will [not] correlate very well with the learning and understanding that [his] students have accomplished". However, he did acknowledge that these types of tests were here to stay, as people could understand them easily.

Two of the teachers, Ms. Fredericks and Ms. Richardson, rejected the usefulness of standardized test scores, preferring performance assessments for the final determination of student achievement. Ms. Fredericks administered the district-mandated tests to her students but did not find the resulting scores useful in her determination of student success. Ms. Richardson expressed a willingness to use the test scores as an indicator of student success at some time in the future but did not trust the accuracy with which the scores reflected her student's accomplishments. In her words, "the tests are too picky, too specific in the nature of the questions."

Of all the teacher participants in this study Ms. Cornette was the one most open to using standardized test scores for the final assessment of her students' achievement because she understood these tests to be part of the "real world" and felt she should start preparing her students to be successful when taking them. In spite of her willingness to accommodate the district mandate she did not feel that the tests tested what she taught and preferred other assessment strategies. When the standardized test scores did arrive, Ms. Cornette found them to be unhelpful.

Both Mr. Gorman and Ms. Cornette acknowledged the reality of standardized testing. They reasoned that while, over time, performance assessments might gain public approval, standardized tests are currently viewed by many as reliable and valid measures of student achievement. They were prepared to comply with the district plan in order to prepare students to be successful on standardized achievement tests that they might have to take in the future. Neither expressed any interest personally in using the test scores to inform their own teaching. Overall, the teacher participants gave only cursory attention to any use of standardized test scores, in spite of the fact that administering the tests and increasing student scores on the tests was a district mandate. Each teacher met the district guideline of administering the required tests. However, as a group, the teachers were not using standardized tests to inform their teaching or determine the level of understanding that their students had achieved.

The preference for informal information on student achievement

The teacher participants in this study assessed student understanding continually during the instruction process and preferred performance type assessments to paper and pencil tests as final assessments. Typically, teachers used assessment strategies at the beginning of lessons to determine what students already knew about a new subject or what they remembered from previous lessons. They embedded assessment in their instruction to determine if students were understanding as lessons proceeded and used various assessment strategies to determine final achievement of understanding.

All of the teacher participants deliberately included strategies to determine the initial understanding of students as new lessons began. Ms. Springer, for example, used concept maps to find out what her students knew about electricity. She referred to the importance of sharing information, emphasizing that "knowledge should not be kept a secret." Mr. Gorman planned and posed questions about the 'big ideas' of each new unit to find out what students understood at the beginning of the unit and to generate discussion among the students about the new topic. Both Ms. Richardson and Ms. Cornette used daily warm-up activities to begin their classes. Students were frequently asked, in Ms. Richardson's class, to reflect on previous learning or experience and make connections to the new learning. Ms. Cornette asked her students to answer questions directly relating to the lesson taught the day before. All teachers used the student responses to determine the level of understanding that students had at the beginning of the lessons and adjusted their instructional plans accordingly. This strategy is nicely illustrated in the lesson on electricity described earlier, when Ms. Springer posed open ended questions after reading the story about electrons, light and heat.

Each of the teacher participants integrated assessment of student understanding into their instructional strategies. Those practices varied from informal questions during whole class discussions, one-on-one conversations between teachers and students as students were working on individual assignments, discussion of homework assignments, quizzes, and reflective writings. Teachers used information gathered from these sources to inform and adjust their instruction in order to provide opportunities for more students to be successful. For example, in the electricity lesson, Ms. Springer checked students' understanding by asking them to write about their "aha" moments after watching a video. On many occasions, both Mr. Gorman and Ms. Fredericks added specific activities to their planned instruction because feedback from student discussions informed them that students had been struggling with specific concepts. Many of the lessons observed in Ms. Richardson's class ended with a review of the day's learning. These question and answer sessions served to inform her of how thoroughly her students had understood the material presented in her daily lessons.

Given the current political climate, which emphasizes the importance of student success on high stakes, multiple-choice achievement tests, teachers who are striving to maintain constructivist teaching practices are working in an environment that would seem to be counterproductive to their efforts.

All of the teachers stated that they would prefer to interview each student individually in order to assess their final understanding of the concepts taught in class. Ms. Springer referred to the importance of "listening to students on an individual basis." Ms. Fredericks was actually able to schedule her classes so that she could have those interviews at the end of several science units. She used the product of performance-type assessments as the focus of those interviews. In the case of an electricity unit she had her students build circuit boards and explain the differences between parallel and series circuits. If they weren't sure of the concepts to be assessed she worked with them in a one-on-one setting until they understood.

Each of the elementary teachers expressed a strong preference for using performance-based assessment in determining the level of student understanding achieved in their classes. Ms. Springer preferred performance assessments including presentations and projects and believed that students demonstrated understanding when they could apply learning in different situations. In the three units of instruction observed in Mr. Gorman's classroom his final assessment of student understanding was embedded in an analysis of a final project. In one case the project involved wiring a model house with series and parallel circuits. Another was constructing and describing the interrelationships in a freshwater ecosystem and a third involved building and flying paper airplanes and explaining how and why variations in design affected the flight of the planes.

The middle and high school teachers also expressed a preference for performance-based assessment. These teachers expressed concern that the district mandated multiple-choice tests did not correspond to the content taught in the classroom and could never accurately test what was learned. There was a clear disconnection between the goals of the teachers in this study in assessing student performance and the district goal of being the number one district in the nation as evidenced by standardized test scores.

During the study the teacher participants implemented assessment strategies that correlated with the suggestions found in the literature on teaching for understanding. The NSES advocate the planning of assessment as an integral part of instruction. According to the NSES, assessments embedded in the curriculum can serve at least three purposes: "to determine the students' initial understanding and abilities, to monitor student progress, and to collect information to grade student achievement" (NRC, 1996, p. 87). These teachers used assessment for all three of the NSES identified purposes. As illustrated in the vignette of Ms. Springer's teaching, teachers made instructional decisions based on initial diagnosis of student understanding and ongoing analysis of student achievement.

The capacity to manage a dilemma of pedagogical choices

Each decision about teaching and assessment required judgments on the part of the teacher about purpose and implementation. The teachers in this study were faced with a dilemma of pedagogical choices. One such dilemma was in determining the level of understanding of science content that their students achieved during and as a result of instruction. When lessons were more heavily packed with content information in the elementary classrooms there was evidence that teachers accepted recitation for understanding. For example, in Ms. Springer's multi-age classroom, there were students who memorized and copied representations of atomic structure as their teacher drew them on the board. These efforts were accepted by the teacher as evidence of understanding.

When lessons were focused on concepts, teachers at the elementary level often used performance assessments to determine the level of student achievement. Mr. Gorman asked his students to wire a model of a house using both series and parallel circuits. Ms. Fredericks had her students build a circuit board and explain to her how it worked. Mr. Gorman determined the level of understanding of airplane flight that his students achieved by having them construct paper airplanes and analyze why changes in the model resulted in change in flight pattern. Ms. Springer had her students read about food chains and webs and then choose different ecosystems and make drawings of different food webs. Mr. Gorman had his students build mini-ponds in the classroom and watch the feeding process of several organisms at different ranks in the food chain. In each case the lesson focus was on broad concepts and the assessment of understanding was based on student use of the concepts to analyze different situations.

The teachers in this study brought a broad range of knowledge and experience to their science teaching, including knowledge of science content and science content pedagogy. Each teacher made daily decisions about what content to teach, which instructional strategies to use, what level of student understanding of concepts to expect and how to assess student understanding. As they implemented their lesson plans and taught for understanding, they faced a dilemma relating to their individual bases of knowledge and experience. Teachers' knowledge of science content and science content pedagogy as well as their personal knowledge and experience impacted their instructional and assessment decisions.

Conclusions

This research raises some important issues about teaching and assessing for understanding in the context of a complex policy environment. The first issue concerns the importance of recognizing the impact of political pressures on classroom teachers. Organizing a constructivist classroom based on the principles of teaching for understanding and then recognizing student understanding when it occurs is problematic at best for teachers at all grade levels. Given the current political climate, which emphasizes the importance of student success on high stakes, multiple-choice achievement tests, teachers who are striving to maintain constructivist teaching practices are working in an environment that would seem to be counterproductive to their efforts. Those who are interested in what teachers are thinking and doing as an important component of efforts towards educational reform and those who are teachers of teachers need to fully understand the dichotomous role that teachers have to fill.

The second issue is about the importance that teachers place on connecting teaching and assessment. Assessment of student achievement, as a topic of discussion, has become a focus of political rhetoric as well as a concern among science educators. The political push for accountability frequently eclipses the intent of the NSES and the educational practices suggested by constructivist learning theory and the principles of teaching for understanding. This study revealed that teachers, as evidenced by their daily practices, are advocates of assessment strategies that allow students to demonstrate what they know and focus on the content that is most important for students to learn. When there is a correlation between, and integration of, instruction and assessment, teachers can develop sound perceptions of student understanding, at both the formative and summative stages. As a result they can adjust their instruction according to student needs and develop higher order assessment strategies. Other evaluation techniques such as the use of external multiple choice style standardized tests may provide quantifiable standards of achievement but the level of learning that they assess may not be consistent with the depth of conceptual understanding that students have achieved. In addition, results of such tests are often not available to teachers during the instructional process and thus do not serve as a tool to inform instruction of the students that were tested.

Finally, we return more directly to the issue of the accountability dilemma and to teachers' capacity to operate within a complex policy milieu. After tracking these five teachers for the almost a year, what strikes us is how competently the teachers were able to deal with the multiple pressures inherent in this dilemma. The five teachers managed to hold these two accountability requirements in a kind of dialectic tension. They recognized the political reality of being held to account by the school board and the public at large through the administration of standardized tests. They also recognized the central importance of a different kind of practical accountability, to individual students and their parents for building student understanding. We found that teachers were able to *both* separate *and* integrate these two kinds of accountability. They wanted students to do well on the standardized tests *and* to develop their understandings. Their day-to-day practice, however, was motivated primarily by a desire to improve understanding, and in doing so they drew information from a wide variety of sources. These sources were mainly informal but also formal (and potentially including, where feasible, available and useful, standardized test scores). We conclude with the observation that the teachers in this study were able to work in a complex accountability milieu by operating with a pragmatic sense of making efficient and effective use of available resources, and motivated by a desire to develop understanding in the students in their care.

References

Adler, P. A. and P. Adler. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377-392). Thousand Oaks, CA: Sage.

Barton, P., and R. Coley. (1994). *Testing in America's schools* (Policy information report). Princeton: Educational Testing Service.

Brooks, J. and M. Brooks. (1999a). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

Brooks, M. and J.G. Brooks. (1999b). The courage to be a constructivist. *Educational Leadership, 57*(3), 18-24.

Duit, R. and D. F. Treagust. (1995). Students' conceptions and constructivist teaching approaches. In B. J. Fraser & H. J. Walberg (Eds.), *Improving science education* (pp. 46-69). Chicago: University of Chicago Press.

Gentile, J. R. (1988). *Instructional improvement: Summary and analysis of Madeline Hunter's essential elements of instruction and supervision*. Oxford, OH: National Staff Development Council.

Hardy, L. (2004). A nation divided on education in 2003. *Education Digest, 69*(7), 4-11.

Holmes, G. A., and L.C. Leitzel. (1993). Evaluating learning through a constructivist paradigm. *Performance and Instruction, 32*(8), 28-30.

Hunter, M. (1979). Diagnostic teaching. *Elementary School Journal, 80* (1), 41-46.

Llewellyn, D. (2005). *Teaching high school science through inquiry; A case study approach*. Thousand Oaks, CA: Corwin Press.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

No Child Left Behind Act of 2001 (2001) retrieved on Sept 28, 2005 from http://www.ed.gov/policy/elsec/leg/esea02/index.html

Perrone, V. (1998) Why do we need a pedagogy of understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research to practice* (pp. 13-38). San Francisco: Jossey-Bass.

Polkinghorne, D. E. (1995). Narrative configuration in qualitative analysis. In J. A. Hatch & R. Wisniewski (Eds.), *Life history and narrative* (pp. 5-21). London: Falmer.

Prawat, R. S. (1989). Teaching for understanding: Three key attributes. *Teaching and Teacher Education, 5*(4), 315-328.

Ryan, J. M. and J.R. Miyasaka. (1995). Current practices in testing and assessment: What is driving the changes? *NASSP Bulletin, 79*(573), 1-10.

Stake, R.E. (1994). Case studies. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). Thousand Oaks, CA: Sage.

Tobin, K. and D. Tippins. (1993). Constructivism as a referent for teaching and learning. In K. Tobin (Ed.), *The practice of constructivism in science education* (pp. 3-21). Mahwah, NJ: Lawrence Erlbaum.

Tomlinson, C.A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wiggins, G.P. and J. McTighe. (2005). *Understanding by design (Expanded 2nd edition)*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wilder, M. and P. Shuttleworth. (2005). Cell inquiry: A 5E learning cycle lesson. *Science Activities, 41*(4), 37-43.

Windschitl, M. (1999). The challenges of sustaining a constructivist classroom culture. *Phi Delta Kappan, 80*(10), 751-755.

Windschitl, M. (2000). The diffusion and appropriation of ideas in the science classroom: Developing a taxonomy of events occurring between groups of learners. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 24 -28, 2000.

Yager, R. E. (1996). Science teacher preparation as a part of systemic reform in the United States. In J. Rhoton & P. Bowers (Eds.), *Issues in science education* (pp. 24-33). Washington, DC: National Science Teachers Association.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). London: Sage.

Adrienne T. Gibson is currently part of the adjunct faculty of Capella University in Minneapolis, Minnesota. She has been involved in public education as a teacher, administrator and curriculum developer for over 32 years. Her current research interests focus on curriculum development and instructional and assessment practices. Correspondence concerning this article may be sent to adrienne_gibson@msn.com

John Wallace is Professor of Science Education at Curtin University of Technology in Perth, Australia. He is currently on leave with a professorial appointment to the Ontario Institute for Studies in Education of the University of Toronto. His research interests include teacher learning, curriculum integration, teacher leadership and school reform.

Kenneth W. Miller, David M. Davison

What Makes a Secondary School Science and/or Mathematics Teacher "Highly Qualified?"

The authors examine the impact of the "highly qualified" teacher aspect of the NCLB legislation. They conclude that a highly qualified teacher requires much more than just content knowledge.

Introduction

Issues facing schools abound from NCLB legislation. In science and mathematics teacher pre-service and in-service preparation, these issues are paramount for institutions of higher education, especially in terms of the definition of "highly qualified" teacher. Within this paper we will discuss the science and mathematics issues facing our state and nation. Specifically we will examine the literature supporting the major / minor teaching certification and NCLB's effect on a rural western state.

The No Child Left Behind (NCLB) act has, without a doubt, been the most controversial federal law of the century. As public schools set about abiding by this law, universities are faced with the task of preparing teachers that are deemed "highly qualified" according to federal definition. While the law specifically evaluates and assesses the public education sector, ramifications of the law thread their way into higher education. Colleges of Education focused heavily on the preparation of the nation's teachers are remiss if they continue to operate as before NCLB guidelines on preparation of "highly qualified" teachers. Specifically in science and mathematics teacher preservice and inservice preparation, these issues are paramount for institutions of higher education, especially in terms of the definition of "highly qualified" teacher.

There has also been general recognition, stimulated in large measure by the Third International Mathematics and Science Study (TIMSS), that the United States has fallen behind other countries in the effectiveness of its science and mathematics instruction. Educational analysts, together with calls for reform from the public, have proposed several solutions to the problems (see below). In essence, universities are faced with the daunting task of preparing secondary teachers with a certifiable major in all teaching fields, a task specifically problematic at the middle and high school levels.

These findings suggest that increased content knowledge, while important, will not of itself guarantee that a teacher will be better able to increase student performance.

Where a teaching minor was once a certifiable option, NCLB legislation now requires states to individualize their certification requirements with a HOUSSE (High Objective State Standard of Evaluation) plan that all but eliminates transferability across state borders.

Whether or not a teacher is "highly qualified" as defined by the NCLB federal education law is not an easy question to answer. To confound the guidelines, each state is asked to further clarify the meaning of several of these requirements, particularly the HOUSSE. Without exception, all states must prepare to meet the 2005-2006 "highly qualified" mandate. There is, however, new flexibility for rural states. Accordingly, the U.S. Department of Education (USDOE) suggests that, "Under this new policy, teachers in eligible, rural districts who are highly qualified in at least one subject will have three years to become highly qualified in the additional subjects they teach. They must also be provided professional development, intense supervision or structured mentoring to become highly qualified in those additional subjects" (2005, p 3). The USDOE also suggests that science teachers need more flexibility in the interpretation of the law. While there is no ambivalence regarding the highly qualified status for mathematics teachers, science teachers must have a major in each science discipline or have a broadfield major. The USDOE suggests, “Now, states may determine—based on their current certification requirements—to allow science teachers to demonstrate that they are highly qualified either in “broad field” science or individual fields of science (such as physics, biology or chemistry)” (p. 3). However, the USDOE must approve the state’s recommendation.

The USDOE is quite specific regarding the definition of “highly qualified”. Teachers must have an academic major in the subject that they are teaching; coursework equivalent to an undergraduate academic major; a graduate degree in the subject being taught; a professional license in the subject being taught; or National Board Certification in the subject being taught. The No Child Left Behind law does not require current teachers to return to school or get a degree in every subject they teach to demonstrate that they are “highly qualified”. The law allows them to provide an alternative method (HOUSSE) for experienced teachers to demonstrate subject-matter competency that recognizes, among other things, the experience, expertise, and professional training garnered over time in the teaching profession. However, without a HOUSSE in action, practicing teachers who are qualified in one subject; e.g., biology, and are also teaching in a minor area like chemistry, have until June 30, 2007 to become highly qualified in these additional subject areas.

Individual HOUSSE plans are as unique as are the individual states. However, one example might serve to help in our understanding of the nature of HOUSSE. Consider a HOUSSE plan that consists of a measurement of content knowledge through the PRAXIS II (4.0 points) added to a GPA measurement (4.0 points) added to an assessment of teaching (4.0 points) for 12 total possible points. States could determine in their HOUSSE plan a cut-off score, for example 9.0 and teachers scoring at or above that score would be deemed highly qualified. Additional scoring component criteria might include a score for professional development and/or teaching experience. Essentially, the formula might look like Figure 1.

Figure 1

Content Knowledge (i.e., Praxis II) 4.0	+	Content GPA (i.e., transcripts) 4.0	+	Assessment of Teaching (i.e. Student Teaching) 4.0	=	Highly Qualified >9.0

Thus, this example of a HOUSSE formula could determine the state’s definition of a highly qualified teacher. But, is it just a means to an end? That is, does even this adjusted definition of “highly qualified teacher” guarantee that our nation’s students will be better served?

Are highly qualified teachers better teachers?

It is clear from the pronouncements from the USDOE discussed above that it can be assumed as unarguable that highly qualified teachers are superior to those who are less well qualified. However, the research on this link between teachers’ background knowledge and their students’ achievement is at best only mildly positive. For example, Ball (1990) found that prospective secondary mathematics teachers had only a cursory understanding of the concepts underlying elementary mathematics. In general, there appears to be no association between the number of advanced mathematics courses a teacher takes and how well his/her students achieve in mathematics (Monk, 1994). This is not a recent revelation. Begle (1979) concluded

> It is widely believed that the more a teacher knows about his subject matter, the more effective he will be as a teacher. The empirical literature suggests that this belief needs drastic modification and in fact suggests that once a teacher reaches a certain level of understanding of the subject matter, then further understanding contributes nothing to student achievement. (p. 51)

Notwithstanding the lack of clear evidence supporting the posited relationship between teachers’ mathematical knowledge and student achievement, there is strong intuitive support for the notion that student achievement is influenced by teachers’ background knowledge. One issue of concern here is the relevance of the science and mathematics courses taken by teacher candidates for increasing understanding of the nature of science and mathematics. In both disciplines, the study of advanced courses takes a prospective teacher to a greater depth in the discipline without ensuring that he/she has a strong conceptual grasp of the foundational ideas and is able to articulate them. This problem was highlighted by Liping Ma (1999) in her comparison of Chinese and U.S. teachers in their handling of routine topics in elementary mathematics. She concluded that the Chinese teachers, even though they had less formal instruction in mathematics, had more profound knowledge of basic mathematics and worked harder at developing effective ways to teach skills. Ma found that U.S. teachers have completed more coursework in mathematics but have less in-depth knowledge of mathematical procedures as evidenced by their responses to fundamental mathematical questions. This result raises the issue of whether a teacher who lacks a deep understanding of mathematics can teach for understanding.

These findings suggest that increased content knowledge, while important, will not of itself guarantee that a teacher will be better able to increase student performance. What is needed for prospective teachers is coursework that focuses on the foundations of the disciplines rather than on studying them to greater depths. The question posed at the beginning of the section "Are highly qualified teachers better teachers?" may thus be answered: Not necessarily. What is called for is a determination of those attributes that do enhance student performance. That is, what are the factors that contribute to teacher quality?

Teacher Effectiveness and Student Learning

The framers of the NCLB legislation have a simple answer to the $64 question: A highly qualified teacher is a more effective teacher. The discussion above suggests that the answer to the question is not so simple. As indicated in the previous section, merely having content knowledge is not enough.

Darling-Hammond and Sykes (2003), in endorsing the assertions above, claim that if effective teaching cannot be associated with improved student learning, then policy attention should be turned to other factors thought to exert greater influence on learning. Further, in their review of the research they find that student achievement is affected more by the teacher than by other factors such as class size or composition.

Thus educational researchers have found that teacher dispositions like collegiality, self-reflection, collaborative and interactive skills, and the ability to adjust personal and professional practice based on reflection are important characteristics of good teachers.

With so much confounding of the significant variables, it is extremely difficult to offer a clear recommendation regarding the most effective way to raise student performance. Since the enactment of NCLB, educational researchers have been addressing this issue, partly in an effort to counter the USDOE assertion that higher content credentials make for a better teacher.

Weiss and associates (2001) at Horizon Research observed many classrooms, rating 59% of them as low in quality, while only 15% were of high quality. They found that teaching strategy–traditional or constructivist–had no influence on whether or not a classroom was high quality, nor did the number of science and/or mathematics classes taken, but that the teacher's preparedness, commitment, and enthusiasm were critical variables.

Emerick, Hirsch, and Berry (2003) of the Southeast Center for Teaching Quality found that NCLB's narrow emphasis on content knowledge has led to lower standards for teachers. They concluded that content knowledge alone does not justify the designation of highly qualified teacher, but that the successful teacher demonstrates understanding of the nature of student learning, the use of multiple forms of assessment, and the ability to differentiate instruction. In short, the high quality teacher will possess appropriate content knowledge, and will also possess considerable background in communicating effectively to students.

Thus educational researchers have found that teacher dispositions like collegiality, self-reflection, collaborative and interactive skills, and the ability to adjust personal and professional practice based on reflection are important characteristics of good teachers. There is little evidence that scores on teacher licensure tests or emergency provisional certification have any impact at all on student learning or measured achievement (Rice, 2003). Still, policy-makers fail to deal with the benefit of existing research or, in this case, lack of research on teacher quality. Certainly research should play a role in policy making decisions.

Administrators and teacher evaluators have long known that simple mastery of the content in science and mathematics by a teacher is not enough. Consequently, teacher evaluation instruments have typically included multiple measures that have been shown to improve student achievement. Teacher quality is not just the number of science courses or mathematics courses a teacher possesses. John Glenn (2001) claimed: "The basic teaching style in too many mathematics and science classrooms today remains essentially what it was two generations ago. By contrast, teaching innovation and higher student performances are well documented in other countries, where students' improvements are anchored to an insistence on strong professional development on teachers" (p. 20). Accordingly, administrators budget for the professional development of teachers with a focus on a strong pedagogical base. It is in this way that they believe they can enhance student learning and achievement.

Where do we go from here?

Throughout the nation many current middle school teachers of science and mathematics will fail to meet the "highly qualified" test by spring 2006. While there has been a reprieve of one year for teachers presently in those positions, that is of little help to the many elementary certified teachers whose academic preparation may consist of no more than a concentration in science and/or mathematics. Those states which offer middle level certification do have an appropriate path for their teachers to follow, and those that have developed HOUSSE plans can provide their teachers with alternative routes to becoming "highly qualified". In states like Montana, where the Office of Public Instruction (OPI) deems a teacher to be highly qualified if teaching in one's area of certification (grades 7 and 8 are regarded as 'elementary'), a head-on collision with the federal mandate is in process. To this point, deans and faculty of the Colleges of Education have lobbied, without success, for compliance with the federal mandate. Efforts to develop alternative routes to the completion of the equivalent of a major in a science or mathematics discipline have run aground where faculty see themselves expected to do more with ever diminishing resources. In our judgment, disaster can be averted only if the state pursues the HOUSSE route.

At the same time, there needs to be a significant body of research on the effects of a teacher's academic preparation on the achievement of his/her students. To suggest that students learn more from a "highly qualified" teacher begs the question posed by this paper. We would all agree that students will do less successful when a teacher is inadequately qualified, but the issue of "how much is enough" in terms of a teacher's science and mathematics credentials is not settled. The position of the USDOE has been made clear in the NCLB law and subsequent implementation statements, but the evidence presented in this paper disputes that conclusion. Our goal is to have every classroom staffed by a teacher with sufficient command of science and/or mathematics to help his/her students to achieve the intended levels of success. We need solid research evidence, not political jockeying or hortatory assertions, to help us determine the most appropriate academic background for the teachers of our nation's science and mathematics students.

References

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal, 90*, 449-466.

Begle, E. G. (1979). Critical variables in mathematics education: Findings from a survey of empirical literature. Washington, DC: Mathematical Association of America /National Council of Teachers of Mathematics.

Darling-Hammond, L., and G. Sykes. (2003). Meeting the "highly qualified teacher challenge *Teacher Education and Practice, 16*(4), 331-354.

Emerick, S., Hirsch, E., and B. Berry. (2003). Unfulfilled promise: Ensuring high quality teachers for our nation's students. Retrieved from www.teachingquality.org.

Glenn, J. (2001). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science teaching for the 21st century*. National Commission on Mathematics and Science Teaching.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Rice, J. K. (2003). Teacher quality: Understanding the effectiveness of teacher attributes. Washington, DC: Economic Policy Institute.

United States Department of Education. (2005). New No Child Left Behind flexibility: Highly qualified teachers. Retrieved 10/25/05 from http://www.ed.gov/nclb/methods/teachers/hqtflexibility.html

Weis, I. R., E. R. Banilower, K. C. McMahon, and P. S. Smith. (2001). *Report of the 2000 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Kenneth W. Miller is Professor of Science Education, Department of Curriculum and Instruction, Montana State University-Billings, Billings, Montana. Correspondence concerning this article may be sent to KMiller@msubillings.edu.

David M. Davison is Professor of Mathematics Education, Department of Curriculum and Instruction, Montana State University-Billings, Montana.