Using Data to "Make a Case" for Mathematics Reform within a K-12 District

September 30, 2005

Ben Sayler*
Black Hills State University

June Apaza
Technology & Innovation in Education

Maggie Austin
Technology & Innovation in Education

*Corresponding author may be contacted at the Center for the Advancement of Mathematics and Science Education, Black Hills State University, USB 9005, Spearfish, SD 57799; bensayler@bhsu.edu.

Abstract

Project PRIME, a targeted Math and Science Partnership focusing on K-12 mathematics within the Rapid City, SD school district, has made extensive use of data to provide key stakeholders with a sense of progress to date and to emphasize areas in need of additional attention. This paper highlights the most compelling findings thus far and includes discussion of the venues and processes used for sharing these data. Preliminary findings have been generated through analysis of 1) student performance on the state's multiple-choice test, 2) student performance on a free-response test developed by the Mathematics Assessment Resource Service, 3) classroom observation ratings, 4) student course-taking patterns, and 5) drop-out rates. Special attention has been paid to gaps and patterns associated with Native American versus non-Native American students. Reduction of race-related disparities represents a primary emphasis for the project. Responses of key stakeholders to the data and lessons-learned about using data to motivate and support reform are also addressed.

Using Data to "Make a Case" for Mathematics Reform within a K-12 District

Project Context

PRIME is an NSF-funded, cohort 1, Math and Science Partnership involving four institutions: Rapid City Area Schools (the district), Black Hills State University (the university), TIE (a regional education service agency), and Inverness Research Associates (the external evaluator). A primary focus of the partnership is to improve the teaching and learning of K-12 mathematics within the district. Sub-goals include reducing the achievement gap between Native American and non-Native American students and improving graduation rates.

The district includes 15 elementary schools, 5 middle schools, and 3 high schools, it employs approximately 500 teachers of mathematics (including elementary and special education teachers), and it has a K-12 enrollment of about 12,000 students. Using grade four as a point of reference, 36% of those students typically qualify for free or reduced-price lunch, 22% are non-white (17% Native American, 5% other races). Rapid City represents the largest off-reservation population of Native American students in South Dakota.

The project's approach to improving K-12 mathematics is essentially two-fold: 1) professional development for teachers (an average of 100 hours of professional development per teacher over 5 years); and 2) implementation of inquiry-oriented, NCTM standards-based instructional materials. Professional development opportunities include a combination of district-wide and building-based offerings ranging from one-day workshops to semester-long graduate classes to study groups to classroom coaching.

Sources of Data

Professional development hours are tracked using an on-line database maintained by the university partner. Elementary teacher leaders, secondary math coaches, project leadership team members, and other facilitators of professional development log the types, dates, and duration of interactions with teachers, principals, parents, and other stakeholders. This database is password-protected, but otherwise accessible and query-able on-line.

Random classroom observations are conducted each year by the external evaluation team, and results are shared with project leadership, devoid of any features that could identify particular teachers. Observations are quantified using the Horizon Classroom Observation Protocol (Horizon Research, Inc., Chapel Hill, NC).

The district administers a state-mandated, multiple-choice student achievement test (Dakota State Test of Educational Progress or DSTEP) each spring at grades 3 through 8 and 11. The DSTEP is aligned to the state math content standards. Math scale scores as well as proficiency designations for each student are made available to the project without names, but with student identification numbers that are assigned by the state. Supplementary student achievement data have been collected by the project over the past two years for a sample of 4th, 8th, and 12th grade students using CTB-McGraw Hill Balanced Assessments in Mathematics (developed by the Mathematics Assessment Resource Service and referred to hereafter as MARS). MARS tests are aligned to the National Council of Teachers of Mathematics (NCTM) standards and are free-response, as opposed to multiple-choice.

Student-level demographics, course-taking data, and end-of-course grades are also supplied to the project by student ID number each year. Finally, interviews of district personnel have been conducted in order to better understand the meaning of course titles from one building to another and to ascertain which courses should be designated as college-preparatory.

Methodology

Student data are linked from one source to another using the state-assigned student identification numbers. The project also links students to teachers using project-assigned teacher identification numbers. Confidentiality issues and technological hurdles associated with gathering all of the desired data and in a readable format have proven challenging but manageable. Cleaning and integrating the data have also proven to be challenging, but again, manageable. Student transience and dual enrollments between buildings are representative of the types of hurdles that must be overcome to assure that every student gets counted, but only once.

Issues of data collection and analysis are the primary focus of the project's internal data team, which generally meets once per month. Additional information about methodology is shared in conjunction with specific analyses described below.

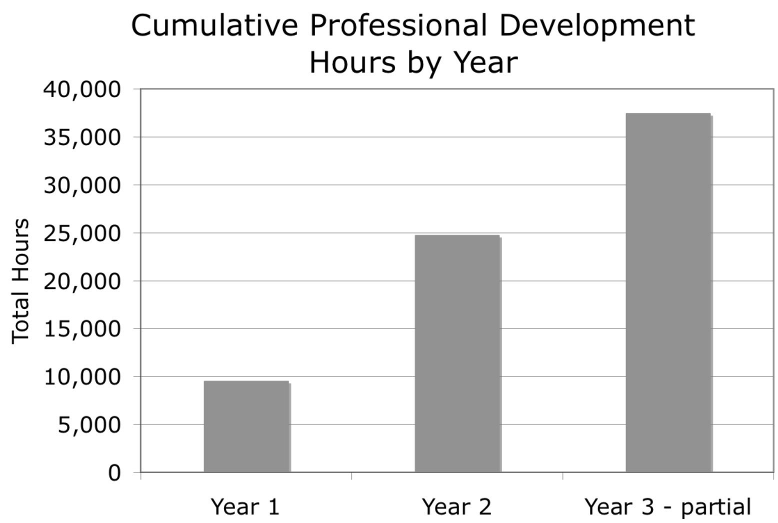
Audiences for the Data

Throughout the project, but especially over the past year (Year 3 of 5), project leaders have made extensive use of data to provide key stakeholders with a sense of progress to date and to emphasize areas in need of additional attention. What follows are examples of graphical representations drawn from multiple meetings over the past year for a variety of audiences: 1) the project's leadership team (which meets monthly and includes three to four representatives from each partner institution); 2) a larger steering committee (which meets twice per year and includes the project leadership team, but also additional administrators from the district's central office, university administrators, K-12 principals, school board members, and elementary math teacher leaders, among others); 3) a group of secondary principals; and 4) the district's secondary math coaches. A few of the findings below were also shared in a district-wide in-service for all instructional staff in the fall of Year 3 (2003-2004) soon after school was underway.

Professional Development Patterns

The aggregate number of professional development hours accomplished by the project serves as a general gauge of overall activity. At the outset, the project set a target of 40,000 hours of professional development over five years (100 hours per teacher for 400 teachers). As can be seen in Figure 1, the project is well on its way to meeting and exceeding that target. This type of data is useful to share to begin a presentation about the project. It tells a positive story, demonstrates a trajectory towards exceeding a project goal, and provides a general sense of the magnitude of the work.

Figure 1



A more detailed look at professional development (pd) hours by trimester (spring, summer, and fall of each year) provides an interesting window on the project's evolution (Figure 2). The project began in the fall of 2002, but it took until the spring of 2003 for a team of elementary math teacher leaders to be hired and for the project to begin providing professional development beyond general awareness sessions. That following summer was a time of high intensity with numerous district-wide institutes. Building-based professional development during the academic year was slow at first when the elementary teacher leaders were new to their positions (and before anyone was hired at the secondary level), but by spring of 2004, teacher leaders were highly active in their buildings and providing significant hours through classroom coaching and facilitation of on-site sessions. The transition from more district-wide workshops to more building-based professional development is especially evident in Figure 3, which displays the change in the distribution of professional development hours by category from year one to two.

Figure 2

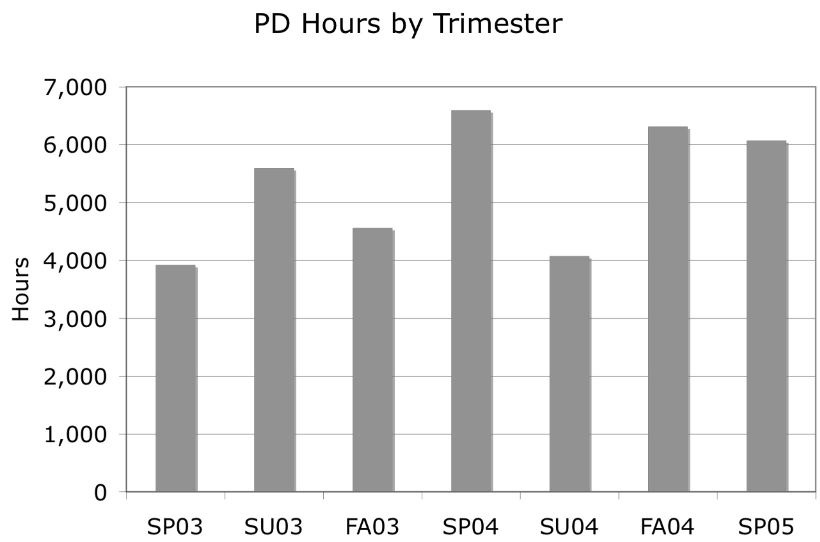
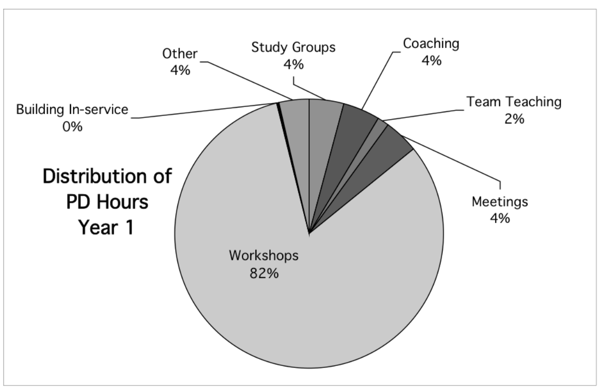
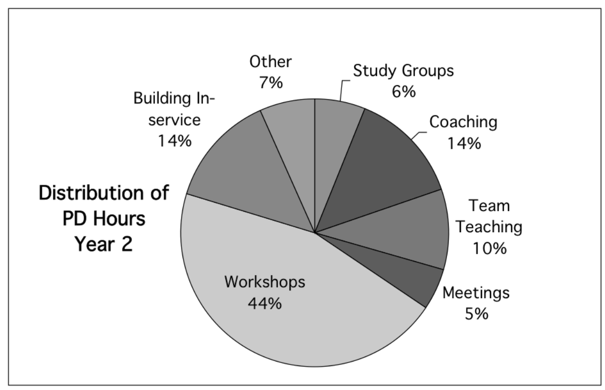


Figure 3



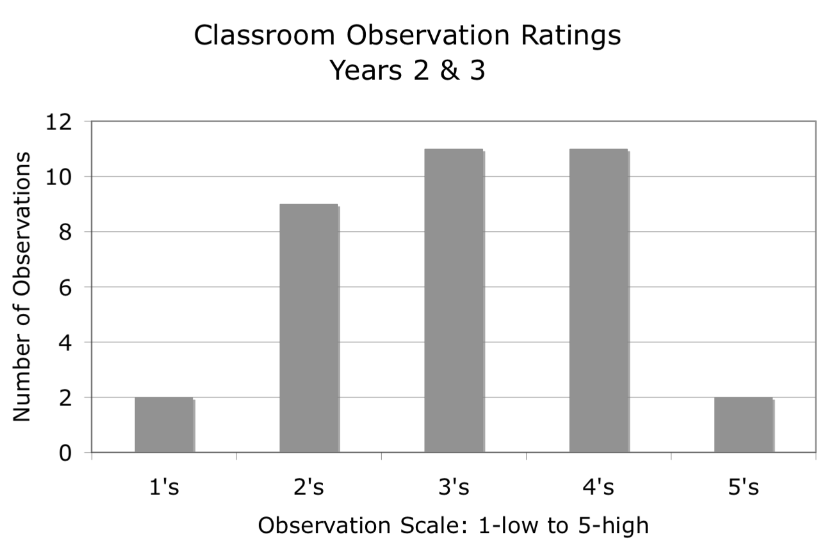


Classroom Observations

While the tracking of professional development hours is important and can help to tell a part of the project's story, merely accruing hours is by no means the ultimate goal. Rather, the project seeks to improve classroom instruction and student experiences. In each of years two and three of the project, the external evaluation team has conducted random observations of math lessons taught by teachers having participated in at least 20 hours of professional development within the project. The distribution of "capsule ratings" over the two years using the Horizon Classroom Observation Protocol are shown in Figure 4. A capsule rating of five on the Horizon scale represents truly exemplary instruction, and a one on that scale represents either "Activity for Activity's Sake" or "Passive Learning." The project has an explicit goal of increasing the percentage of lessons rated at three or above. As can be seen in Figure 4, the majority of lessons observed to date have been rated at three or above, but clearly there remains room to improve.

Teachers who are observed receive no feedback from the external evaluator observing them. It is therefore especially important for the project to share this type of data with teachers, albeit in aggregate form, to demonstrate that observation data are actually being put to use and to emphasize the project's commitment to teacher anonymity associated with these observations.

Figure 4

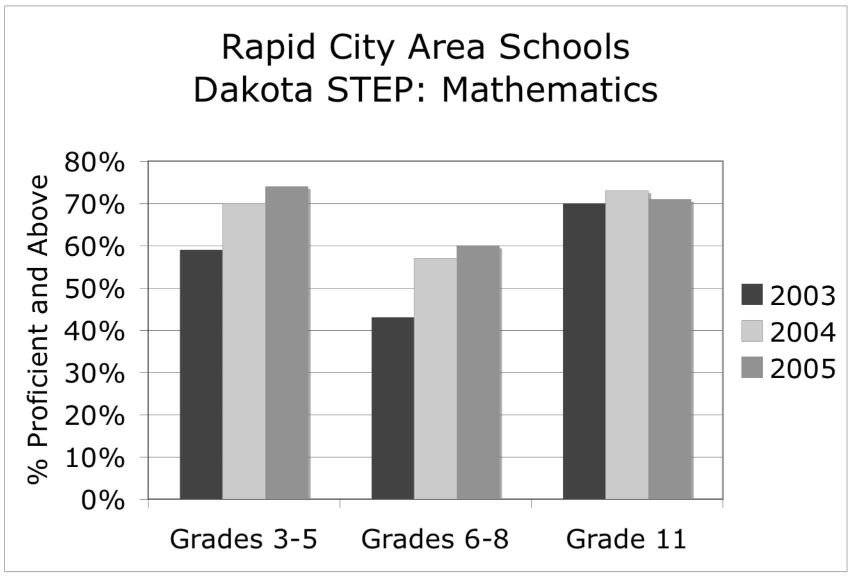


Student Achievement and District Enrollment

The project views high quality lessons as an intermediate indicator along a path toward improved student attitudes about mathematics, increased numbers taking upper level courses, improved achievement, and decreased achievement gaps. The first piece of data related to student achievement is one of general success. As is evident in Figure 5, the majority of students in Rapid City were proficient or advanced as measured by the state test in 2005, and at elementary and middle school grades, there is indication of steady improvement over the three years of the project. Although these data provide no information about causality, it is interesting to note that NCTM standards-based instructional materials have been phased in at the elementary and middle grades over these same years, and implementation has been increasing. The specific materials that have been adopted are *Investigations*, developed by TERC and published by Scott-Foresman, at elementary grades and MathScape, developed by Education Development Center and published by Glencoe/McGraw Hill, at middle school grades. At high school, a shift to more NCTM standards-oriented instructional materials is currently under consideration with pilot-tests underway during the current academic year.

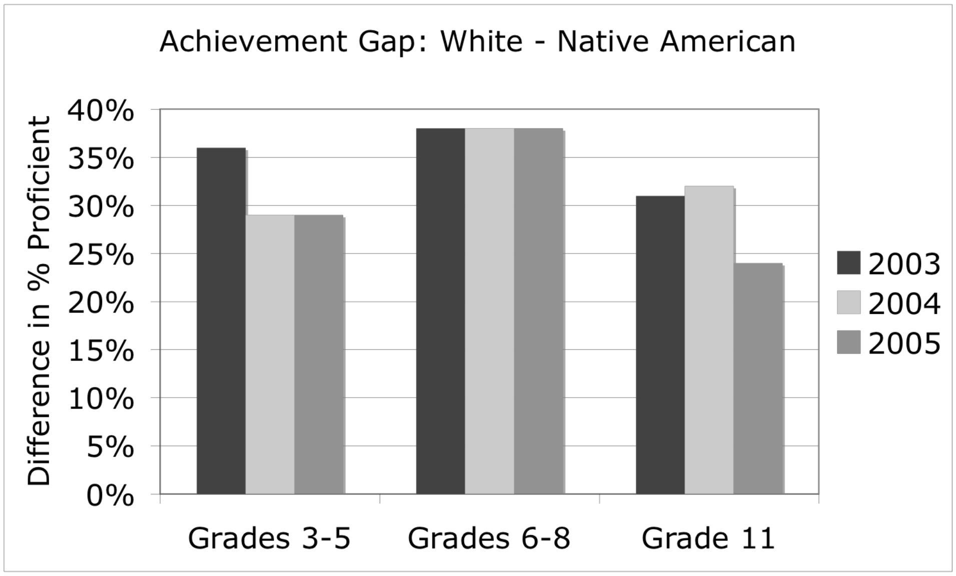
To put the achievement results represented in Figure 5 into context, it is useful to know how achievement scores were changing outside of the district over the same time period. A preliminary analysis of statewide data suggests similar upward trends in the percentages of elementary and middle grade students scoring at proficient and advanced levels. A comparison of actual scale scores (a more sensitive measure than percent proficient) is now in order to gain a better sense of the degree to which project activities might be related to improvements in achievement on the DSTEP. Access to those data is currently being sought.

Figure 5



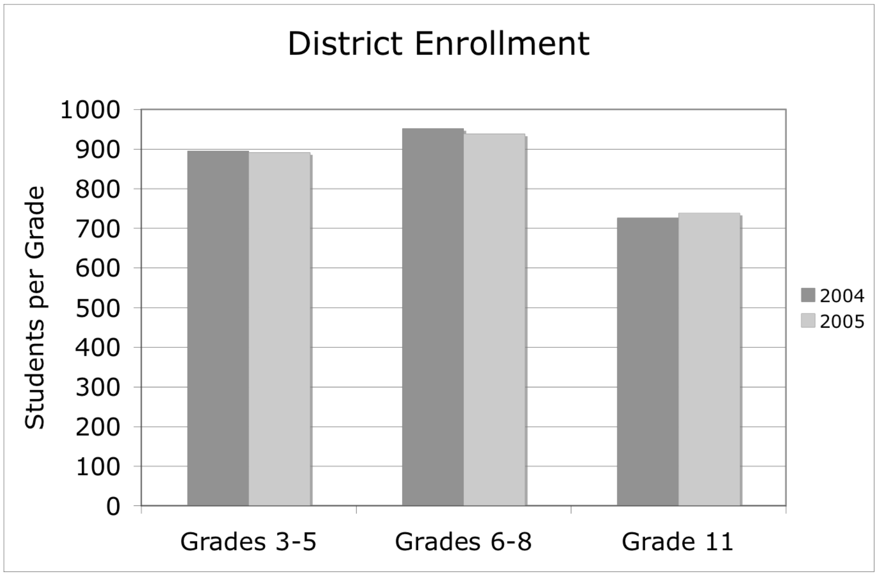
The general degree of proficiency indicated by Figure 5 could potentially lead to complacency. More sobering, however, is the gap in proficiency between white and Native American students that appears in Figure 6. This gap is computed by subtracting the percentage of Native American students who score at the level of proficient or advanced from the percentage of white students who score at those levels. The gaps are distressingly large across the board (consider, for example, that a gap of 30% is consistent with about 70% of white students, but only 40% of Native American students achieving at the proficient level or above). The decrease in the gap at elementary grades from 2003 to 2004 was cause for celebration, but alas, there was virtually no improvement from 2004 to 2005 at those grades. The lower gaps at high school are somewhat comforting at first glance, but ultimately, they appear to be more an artifact of non-proficient Native Americans leaving the system by 11th grade than improved performance.

Figure 6



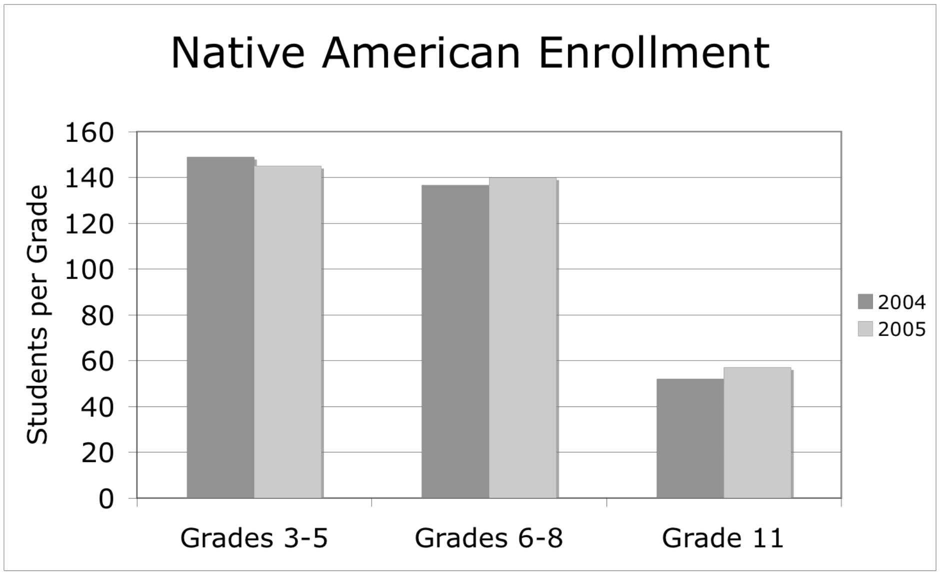
A closer look at district enrollment patterns in Figure 7 shows that indeed, there are notably fewer students in the district at grade 11 than in previous grades. While the 11th grade classes in Figure 7 could have been small to begin with, historical enrollment data (not shown) reveal marked declines in enrollment in individual cohorts between 9th and 10th grades.

Figure 7



Losing a couple of hundred students through the high school grades is regrettable, but it is substantially more concerning when one takes a closer look at precisely which students are leaving. Figure 8 shows that few Native American students, in particular, are making it through 11th grade. Again, there exists a question about whether this specific cohort has had low numbers of Native American students all along, but again, other data indicate that indeed, the number of Native American students graduating is consistently well under half of the number entering high school as freshmen.

Figure 8



Supplemental Measure of Achievement

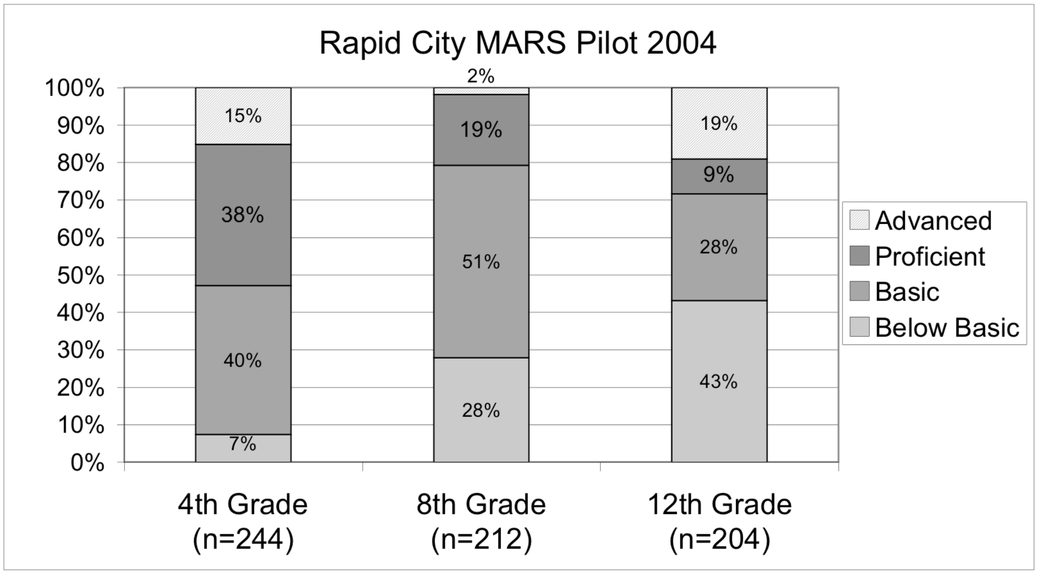
In the spring of 2004, project leaders administered an additional measure of student achievement to supplement the DSTEP. As a pilot, the project selected 11 fourth grade math classrooms and 12 eighth grade math classrooms (one class of each of the eighth grade math teachers in the district) for testing using MARS performance measures. These two grades were selected to represent elementary and middle school, respectively. The motivation behind supplementing the DSTEP with MARS was a speculation that perhaps MARS performance might be more sensitive to and consistent with the NCTM standards-based instructional materials being phased in across the district. It was also thought that using MARS might yield interesting results related to process standards such as communication since, in contrast to the DSTEP, MARS items are free-response.

The project has examined the relationship between classroom instructional materials and student achievement at grade four, in particular, with preliminary results reported elsewhere (Apaza, Sayler, and Austin, MSP Evaluation Summit, 2005).

In early fall of 2004, a collection of non-math 12th grade classes were added to the pilot. The goal was to test a sample of 12th graders that was representative of the entire population in terms of SES, race, and course-taking patterns. The choice to sample 12th grade in the early fall was intentional so that comparisons could be made with 11th grade DSTEP results from the previous spring. The project wanted to ensure that 12th graders who were not currently enrolled in any math class were included in the sample, and that is why non-math classes were used as the sampling unit.

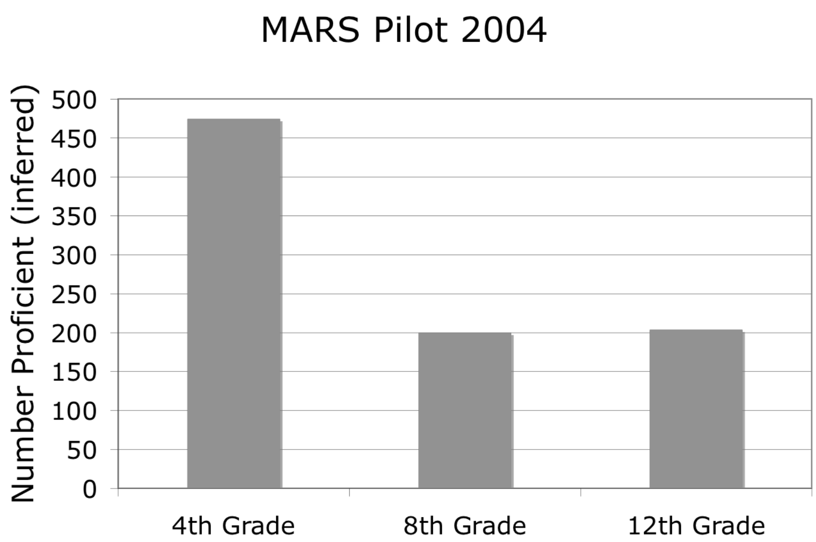
Results of the 2004 MARS pilot appear in Figure 9. The designations of "Advanced," "Proficient," "Basic," and "Below Basic" have been applied to the four proficiency categories that MARS uses. In general, many fewer students scored in the top two categories on MARS than they did on the DSTEP, showing increased room for improvement on MARS compared to the DSTEP. That fact alone could merely be the result of more stringent cut scores, but what are perhaps more illuminating are differences between the grade levels. More than half (53%) of fourth graders were proficient or advanced, while only 21% and 28% were proficient or advanced at grades 8 and 12, respectively. It is certainly conceivable that the older students took the test less seriously than did the elementary students, but even if that is true, those attitudinal differences would be likely to have pertained for the DSTEP as well since neither test had "high stakes" implications for students.

Figure 9



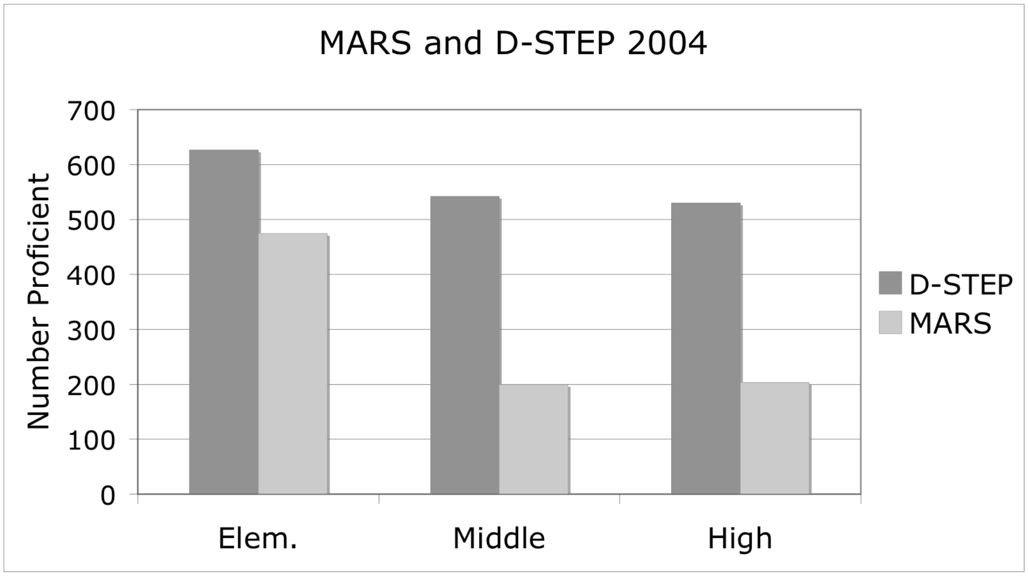
In addition to considering percentages of students in each proficiency category, it is also instructive to look at the absolute number of students who were proficient or advanced. In Figure 10, we have inferred the results of the sample to the entire population at each grade level. Viewed this way, it is evident that more than twice as many fourth graders as eighth graders or twelfth graders were in the top two proficiency categories on MARS. Additionally, it is interesting to note that although the high school students appeared to have outperformed middle school students in Figure 9 above, those grade levels are virtually identical when viewed in terms of absolute number proficient or advanced.

Figure 10



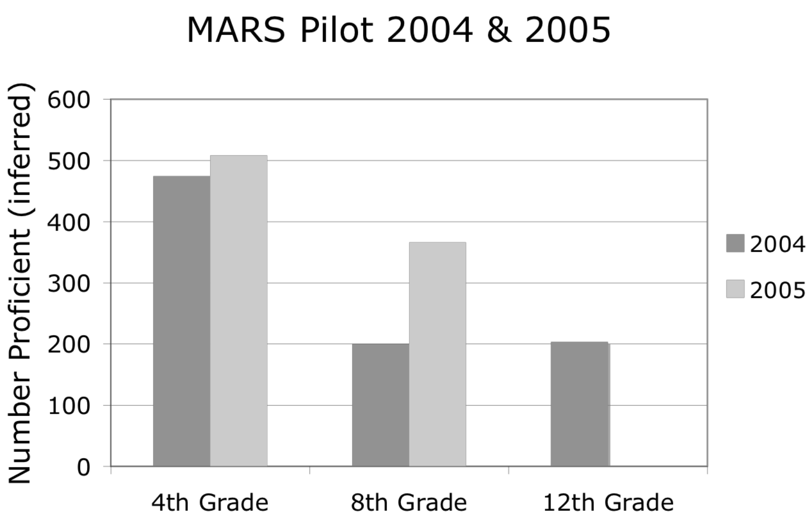
Extending the same inferential analysis to the DSTEP results of 2004, yields an interesting comparison, which is displayed in Figure 11. Here, it is evident that MARS and DSTEP performance is much closer at elementary than it is at middle and high school. A possible interpretation is that in middle and high school, students are spending less time demonstrating how they approach a problem or justifying their answers, both of which are valued by the MARS scoring protocols. The results in Figure 11 are consistent with an argument that implementation of standards-based instructional materials at elementary grades over the past few years is helping students to perform better on MARS when compared to students at the upper grades, where standards-based instructional materials are less prevalent. This line of reasoning deserves additional attention.

Figure 11



Finally in this section, results are shared in Figure 12 of the 2005 continuation of the MARS Pilot. Similar sample sizes and procedures were used for both fourth and eighth grades, and different, but comparable forms of the MARS assessments were administered. Twelfth graders will be tested in the early fall of 2005, so their data do not appear. Gains are evident in Figure 12 in terms of the numbers of students who are proficient at both grades. The gain at grade eight is nearly a doubling. Concurrent with that growth are the facts that middle school teachers have been actively participating in professional development over the past two years and that much of the middle school professional development has focused on the NCTM standards-based materials that are being phased in at those grades. A careful investigation into the *cause* of the eighth grade gain is worth pursuing.

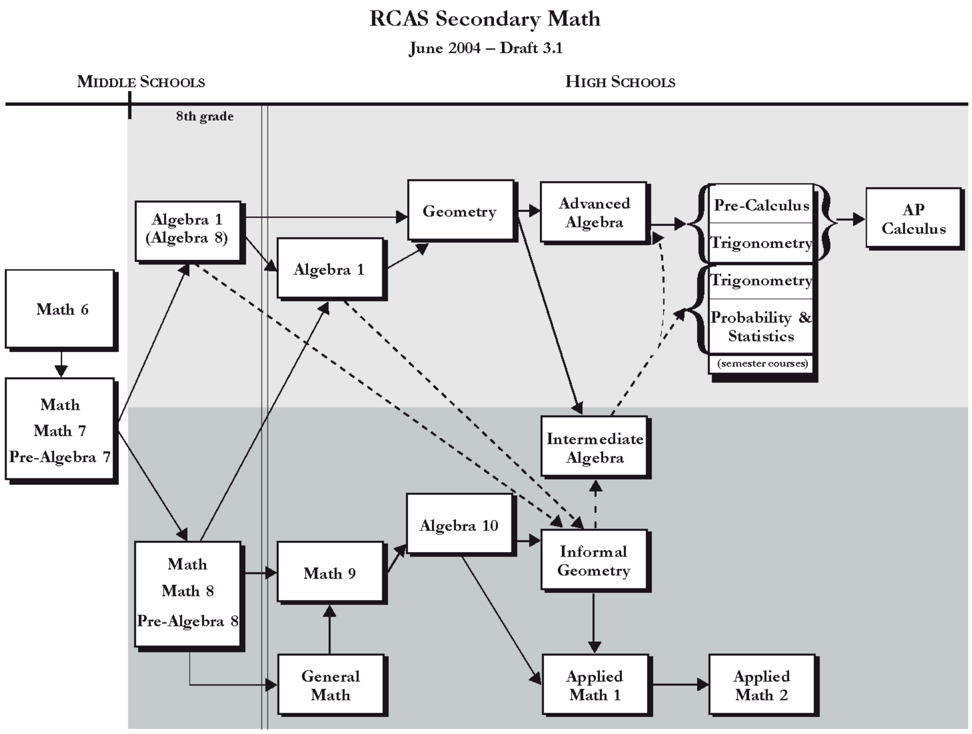
Figure 12



Course-taking Patterns

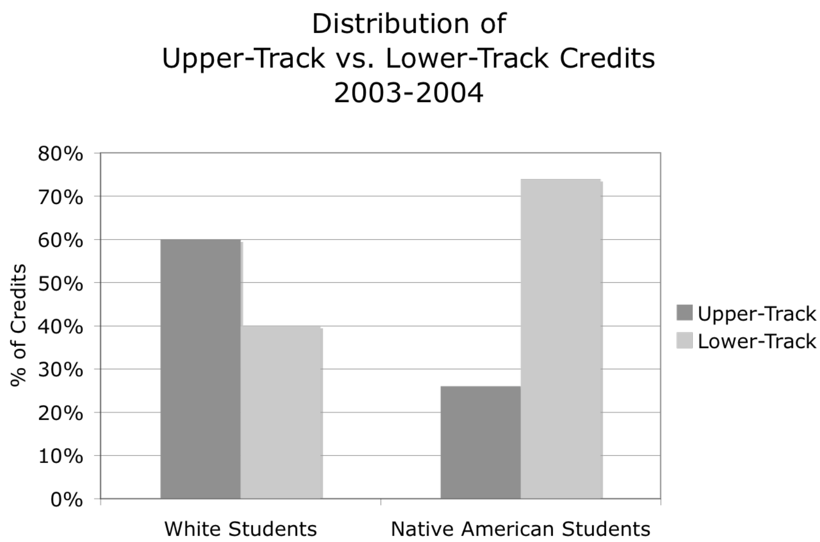
This final section of data pertains to the paths that secondary students typically follow through mathematics coursework. Figure 13 is a map of current high school courses along with a categorization of each as either "upper track" or "lower track." These categories are intended to reflect college-preparatory versus non-college preparatory mathematics across the district's multiple high school buildings. This map represents a considerable investment of time by the project's internal evaluator; the categories were settled upon through a back-and-forth process of interviewing secondary math personnel and sharing drafts of the map until general consensus was achieved.

Figure 13



With the map in place, it became possible to examine numbers of students enrolled in each track. Figure 14 shows a breakdown of credits earned in each track for the two largest racial groups (white and Native American). From this figure, racial disparity is immediately obvious.

Figure 14



In conclusion to this section, it is valuable to examine more closely the absolute numbers of students successfully completing upper-track courses. Raising the number of students pursuing and successfully completing upper-level mathematics courses is an explicit goal of the project, which makes this analysis is especially germane. Data from the 2003-2004 academic year are shown in Figures 15 and 16. Figure 15 represents all students, and Figure 16 focuses solely on Native American students. The "levels" in these figures are those defined by Westat as part of the overall data collection process for all of the Math and Science Partnership projects funded by the National Science Foundation. With regard to Project PRIME, Level 1 translates to Algebra 1, Level 2 translates to Geometry, Level 3 translates to Advanced Algebra, and Level 4 translates to Pre-Calculus and the other courses represented in Figure 13 between Advanced Algebra and AP Calculus.

Figures 15 and 16 both have the same general shape, and both reflect declining numbers of students succeeding as the level increased. It is also true (though not show) that the number of students attempting upper level math classes declined as the level increased. In fact, the passing rate at the two highest levels exceeded 90% of students attempting those classes. The general state of affairs is worthy of documentation and attention and leaves room for improvement, but it is perhaps not too surprising. What reinforces the intensive need for this project, however, and serves as a reminder of the daunting magnitude of the task at hand is the near absence of Native Americans succeeding at the highest levels. Compared to elementary grade-level cohorts of about 140 Native American students, a mere 15 Native Americans succeeded in Advanced Algebra in 2003-2004, and only 3 passed AP Calculus.

Figure 15

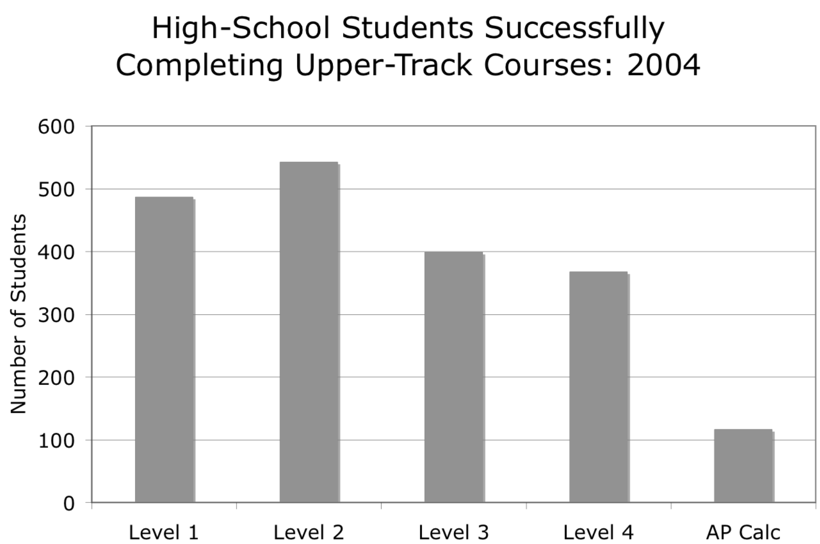
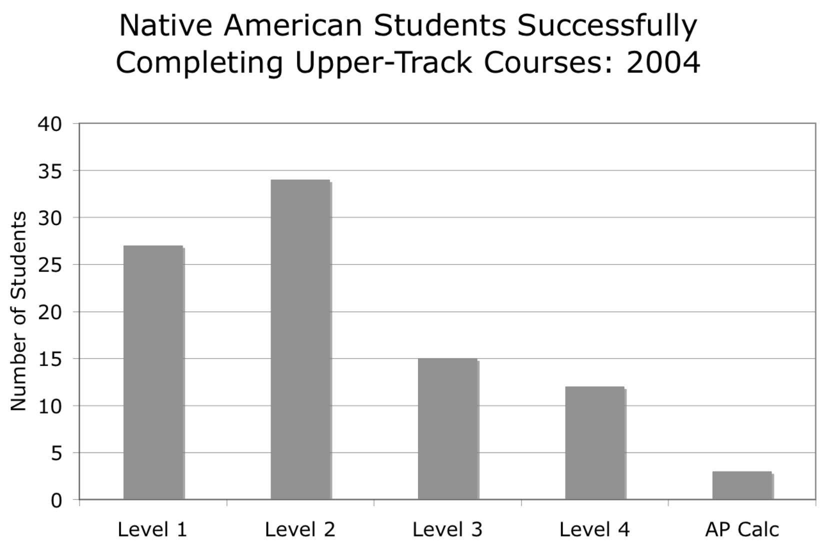


Figure16



Lessons-learned about Sharing Data with Key Stakeholders

The first point is that project data, as presented here, represent significant food for thought. These analyses promote lively discussions and serve as glimpses into interesting research questions, but it is important to recognize that they do not represent formal, completed research *per se*. It is tempting for audience members to speculate from these data about causality, so it is the obligation of the presenter to interject caveats and promote cautious interpretation. In many cases, companion investigations have been undertaken in an effort to rule out possible contaminating factors (e.g., comparing characteristics of samples to overall populations), but again, the data are being shared as works in progress.

Project leaders believe that it is valuable and, in fact, critically important to share emerging results with key stakeholders. Elementary math teacher leaders, for example, benefit from seeing the fruits of their labor logging professional development interactions with teachers. Classroom teachers, as another example, benefit from seeing that classroom observation results are being utilized responsibly and that strict anonymity is being honored.

Project leaders have structured stakeholder meetings not only to share results, but also to give stakeholders a voice in the work. Reflection upon the data is facilitated through small group discussions after several types of data have been presented. Sometimes the small groups are formed such that stakeholders with the same roles meet together; other times these groups are formed across partner institutions or roles to widen the view. Examples of questions posed for small group discussion include:

- What is your initial reaction to the data just presented?
- What questions do these data raise for you?
- What actions do the data suggest for you in your role? For others?
- What advice do you have for project leaders?

Facilitators of these data sessions pay careful attention to issues of format and tools (question/answer sheets to record reactions, for example) to foster rich discussion and input. Both large and small group discussions have yielded important feedback and direction for the project.

Presuming that data such as these are going to be shared as a project is unfolding, they need to be shared thoughtfully and strategically. A persistent barrage of negative results would certainly be damaging to morale, whereas sharing only good news could promote or perpetuate complacency. Considerable care has been taken to strike a productive balance, and stakeholder responses been gratifying thus far with sessions generating probing questions, thoughtful reflections about educational and social values, sensitivity to issues of culture, and what seems to be a sense of camaraderie around common goals.

This project is asking for and spawning major changes. As evidence related to these changes emerges, stakeholders benefit from opportunities to examine it, and they strengthen the project by revisiting its strategies in light of the emerging evidence.

Acknowledgements

The authors are grateful to the many partners within Rapid City Area Schools, TIE, Black Hills State University, and Inverness Research Associates who work in close collaboration to support this endeavor.

Financial support has been provided in part through grant EHR-0227521 from the National Science Foundation.