WORKING
PAPER

Pittsburgh Public Schools' Data Systems

Opportunities for Analysis in Support of Data-Driven Decisionmaking

BRIAN GILL AND JOHN ENGBERG

WR-281-EDU

April 2009

Prepared for the Pittsburgh Public Schools

RAND EDUCATION

PITTSBURGH PUBLIC SCHOOLS' DATA SYSTEMS: OPPORTUNITIES FOR ANALYSIS IN SUPPORT OF DATA-DRIVEN DECISIONMAKING

Brian Gill and John Engberg

RAND

WR-281-EDU

21 July 2005

Overview: Data to support school improvement in Pittsburgh

In the era of high-stakes accountability under the federal No Child Left Behind Act (NCLB), educators are increasingly interested in data and analysis that can effectively inform decisionmaking at the classroom, school, and district levels. Raising proficiency rates for all students, including disadvantaged subgroups, as NCLB mandates, is likely to require tools for diagnosis and evaluation that are much more intensive and sophisticated than public schools have historically had available.

The Pittsburgh Public Schools (PPS) intends to use data and analysis to support a district-wide effort to increase student achievement and reduce the racial gap in achievement. Over the last several years, PPS has made substantial investments in the development of a data system that is designed to create the electronic infrastructure to support data-driven decisionmaking at all levels. The PPS data system (known as RTI, for "Real-Time Information") is designed to link all of the district's information about students together in one place.

PPS commissioned RAND to assess the extent to which the RTI system will permit the use of longitudinal statistical techniques to inform data-driven decisionmaking in the district. To be sure, some kinds of decisions can be directly informed by "raw" data in the RTI system. For example, a teacher who needs to know a particular student's strengths or weaknesses with respect to specific mathematical skills can consult detailed data on the student's prior test results. But well-informed policy-level decisions require sophisticated analysis of data to provide estimates of the *effects* of programs, policies, schools, and teachers—distinguished, as much as possible, from the effects of families, peers, and neighborhoods. In the absence of experiments that randomly assign students, teachers, and schools to differing programs—rarely possible in educational settings—the best way to assess educational effects is with longitudinal modeling techniques (including "value-added assessment," or VAA) that examine achievement and other outcomes for individual students over time, as they move through the school system.[1]

The longitudinal analysis of student data in Pittsburgh has the potential to inform a wide variety of policy and instructional decisions in the district. Currently, the state's testing regime (the PSSA) provides snapshots of the achievement of a few cohorts of students each year—information that may be useful for statewide comparisons and accountability purposes, but is of limited value for informing district decisions about the effectiveness of

[1] To be sure, some methodologists would argue that randomized experiments are the only way to provide valid inferences about causal effects. Although we acknowledge that randomized experiments represent the "gold standard" for causal inference, decisionmakers in the real world usually do not have the luxury of designing experiments to inform every policy decision. Statistical methods that employ longitudinal growth models are the next best thing—usually far superior to other non-experimental methods—and can be used with existing data without imposing an experiment on educators.

programs and policies. By contrast, VAA and other longitudinal modeling methods that examine trajectories of achievement for individual students over time have the potential (1) to provide better information to evaluate the effectiveness of instructional programs and initiatives; and (2) to identify schools and practices that are especially effective at promoting student achievement and attainment and at reducing racial gaps. The Pennsylvania Value-Added Assessment System (PVAAS) takes a step in this direction, using multiple years of achievement data for individual students. But the RTI data system operating in PPS may offer the potential to go well beyond PVAAS by including a wide variety of additional data in longitudinal analyses—thereby offering opportunities to examine effects related to student behavior and graduation rates as well as test scores, to explore the influences of family and neighborhood factors in achievement, and to specifically address factors related to the achievement and attainment of low-income and African-American students.

Whether RTI can be used to conduct the kind of longitudinal modeling analyses that can provide usable estimates of educational effects depends, first of all, on the construction of the data system and the extent to which it is populated with appropriate information. These are the issues that RAND has examined this spring and summer. This report describes RAND's findings about the RTI system's capacity to support VAA and other longitudinal modeling techniques that might inform district decisions aiming to promote student achievement and reduce the racial gap in achievement.

In general, we find that RTI is well-suited to the application of such methods. RTI data are indeed organized in a way to permit longitudinal modeling of student achievement, and consequently to permit analyses that can usefully inform district decisionmaking to promote student achievement. Data from the various assessments conducted in the district are linked to individual students longitudinally over time, meeting the first requirement for the application of longitudinal modeling methods. Moreover, the assessment data are also linked to student-level information about family background, course taking, disciplinary action, enrollment, and attendance. Although the historical completeness of the data varies somewhat with different types of data, most of the key pieces appear to be well-populated for current and recent years.

Two important caveats to these findings deserve notice. First, reliably estimating educational effects requires more than a well-constructed data system and the application of longitudinal statistical techniques. For any particular policy or programmatic question that may be of interest to PPS decisionmakers, the usefulness of longitudinal methods will also depend on several other factors including the numbers and characteristics of program participants and non-participants. These issues will need to be examined for each specific policy/programmatic question that PPS may wish to examine using longitudinal methods.

Second, assessments of growth in student achievement are complicated by the use of varying assessments in different years. The results of VAA, in particular, are sensitive to the particular assessment used. Longitudinal modeling methods are possible even with assessments that differ from grade to grade, but in the absence of a consistent developmental scale across grades, their interpretation requires caution. Nevertheless, although the use of varying tests complicates the assessment of growth, it also creates opportunities to examine the breadth of knowledge and skills mastered by students in the district.

The next section presents a brief overview of the RTI data elements that are likely to be relevant to longitudinal-modeling analyses that can inform district decisionmaking. We then enumerate a variety of specific policy and programmatic questions (many of which were raised in our initial proposal to PPS, and others of which have arisen during conversations with PPS staff during the course of our review) that might be addressed by applying longitudinal modeling methods to RTI data. Some important analyses might be conducted immediately, with existing data in the system, while others should become possible in the near future as additional data are added. The report concludes with recommendations about the integration and collection of additional data in the system, and suggestions for possible first steps for implementing some longitudinal models.

Relevant data components in the PPS/RTI data system

In general, to inform the kinds of policy and programmatic decisions of interest to PPS, longitudinal modeling methods require that longitudinal data on student outcomes be connected to data on inputs such as student demographics, family background, schools, teachers, neighborhoods, and participation in particular educational interventions. We summarize RTI's coverage of these components below.

- *Student outcome* information in RTI includes assessment results for New Standards Reference Exams (NSRE) and PSSA tests back to 1999-2000, for TerraNova math assessments since 2002-03 and reading and writing assessments since 2003-04, for the Balanced Assessment in Mathematics (BAM) (item-level results) since 2001-02, and for the PASS science assessment since 2001-02.[2] (Additional details on the years and grade levels for which particular assessment results are included in RTI are provided in Appendix 4.) We examined the completeness of the data for the mathematics assessments (as described in Appendix 4), and found that RTI in fact includes data for the great majority of enrolled students in recent years. The rate of missing data increases in higher grades, however, which will create some complications for analyses of high-school grades. Other student outcome information in RTI includes grades received in each course taken back to 1999; very detailed misconduct data, including a categorical code and descriptive notes about the incident, links to other students involved, and punishments since 2000; and whether students graduated from PPS. Dropout information is also included, but there is reason for skepticism about whether all dropouts are accurately recorded.[3]

[2] Additional data from some assessments in earlier years can be found in other district databases, but are not currently in RTI.
[3] Some of the withdrawal information on which drop-out status is based might be imperfect. All school districts have difficulty tracking down accurate withdrawal information for dropouts, and they have little incentive for doing so. Some students who are coded as moving out of the district might actually be drop outs.

- *Student demographics* in RTI include age (and date of birth), race, and sex.

- *Family* information in RTI begins with home addresses and contact information for each student. Each student can have multiple contacts. The relationship between a student and each contact has a relationship code that indicates the nature of the relationship (mother, father, etc.) and an indication of whether this contact is the student's guardian. The address of the contact is also listed, so it can be determined whether the student's primary address is the same as the contact. Although RTI does not retain information on previous home addresses and contacts, information in the annually archived version of RTI could be used to assess family mobility and changes in family structure.[4] For all students who apply for free or reduced price lunch, RTI also has information on family monthly income and number of family members. Finally, RTI includes information connecting siblings to each other.

- *School* information recorded in RTI includes the identification of each school attended by each student, back to 1989. In addition, the expected feeder pattern for each currently enrolled student is reported in RTI. RTI permits an examination of how each school organizes its teaching faculty (e.g., whether teachers teach the same classes from year to year, and whether teachers specialize

[4] This information is entered by school personnel based on information sheets provided by the students at the beginning of each year. Only current student/contact links are kept in RTI. When this information is edited, prior information is over-written. Therefore, longitudinal information is only available by retrieving archived versions of RTI from the annual archive.

in particular subjects) back to 2002. For each school, RTI permits examination of what lower level schools and what neighborhoods the children come from.

- *Teacher* information in RTI is, as yet, relatively limited. Students can be connected to current classrooms and current teachers, but RTI does not yet use unique teacher identification codes that would connect teachers to students longitudinally over time. By contrast, the separate PeopleSoft system used by the district for personnel purposes contains many characteristics about teachers. For each teacher, it includes demographics, education and credentialing information, experience in the district and prior to working in the district, areas of certification and areas of assignment, attendance, and workshop attendance.[5] (More details about the teacher characteristics available from PeopleSoft are available in Appendix 1.) PPS information technology staff plan to link the PeopleSoft database to RTI before the end of this summer, with the aim of creating links back to 2002.

- *Neighborhood* data begins with students' home addresses. RTI's storage of student addresses could allow the linking of Decennial Census information about the socio-demographic characteristics of each neighborhood at the census tract or zip code level. Student and family information can also be summarized into neighborhood clusters, to assess the characteristics and academic performance of the other students and families who live near each student.

[5] This is based on an interview with PPS staff who are familiar with the PeopleSoft data system. RAND did not review actual PeopleSoft data.

- *Program participation and implementation* information in RTI is relatively limited at the moment, but the system has the potential to include substantial information about any variety of educational programs and interventions underway in the district. Currently included in RTI for each student are course participation (since 1989), special education status, participation in gifted and magnet programs, and daily attendance (since 2002). Codes for various supplemental programs are now being added to RTI by PPS staff. In addition, the district's mathematics department maintains a separate database of curriculum implementation information (discussed in more detail in Appendix 2).

Possibilities for longitudinal modeling with RTI data

We present here a series of policy- and program-related questions (derived from our original proposal and subsequent discussions with PPS staff) that PPS might want to address using longitudinal modeling methods, and discuss whether the system currently includes the data needed to address those questions. In general, our findings are encouraging: the RTI data system is designed and populated with the kind of data that can be analyzed using longitudinal methods. For any particular question that PPS may wish to examine, the utility of the longitudinal methods will depend largely on the distribution of students across different classes, schools, and educational programs (as noted in the general caveats on p. 5).

1. How are school, student, family and neighborhood characteristics related to the racial achievement gap, and where might efforts to reduce the gap be directed? To what extent is the achievement of African-American students related to

observable socioeconomic factors? Are some schools working effectively to reduce the gap? What is the scope of the racial gap in graduation rate? Is the achievement and attainment of African-American students related to the level of racial integration in schools and academic programs?

RTI currently includes sufficient information to permit the use of VAA and other longitudinal modeling methods to address each of these questions, pending an examination of the distribution of students across schools in the district. The combination of assessment, attainment, demographic, family, school, and geographic information included in RTI creates opportunities for analyses that have the potential to provide better diagnostic information than any previously available on the sources of racial achievement disparities in the Pittsburgh Public Schools. Such analyses might go even further if the RTI data are supplemented with additional US Census information on neighborhoods and additional PPS information on student participation in specific educational programs.

2. What are the effects of attending K-8 schools vs. middle schools? For students who have similar records in fifth grade, how does the eighth-grade achievement compare for those who attend K-8 schools vs. middle schools? How does the high-school achievement compare? Does attending a K-8 school affect the likelihood of dropping out of high school?

All necessary information is available in RTI to analyze issues related to achievement effects of K-8 schools vs. middle schools, provided that the number of comparable students in K-8 schools and middle schools is sufficient. Effects on dropping out will be more difficult to gauge, given questions about the reliability of dropout data. Nevertheless, it may be possible to examine these effects indirectly by assessing the probability of graduating from a PPS school.

3. What are the characteristics of students who are at risk of falling short of proficiency on the PSSA? Can the district identify those students and their specific academic needs early, to allow instructional interventions for those students? Are the district's occasional and annual assessments well aligned with the PSSA?

The RTI system includes sufficient data to address each of these questions. The diagnostic analyses are currently limited only by the length of historical data, which goes back several years for most assessments. A full assessment of the alignment of tests would require intensive examination of individual test items, but a preliminary assessment of alignment might be undertaken immediately, examining student-level correlations in scaled scores or sub-scales across different tests.

4. Is there evidence that school size affects student learning, in the short term or the long term? Does school size affect dropout and graduation rates?

All necessary information is available in RTI to examine the relationship between school size and outcomes. The primary limitation here is the number of schools in the district, which creates some limits in terms of statistical power.

5. Which schools appear to promote the largest gains in student learning? What are the characteristics of those schools and can they be replicated in other schools?

Existing RTI data can be analyzed with longitudinal models to identify which schools promote the largest gains in student achievement. Comparisons among schools will require some caution, however, because student populations may differ substantially, in ways that may not be fully captured by previous test scores or other student measures. Analyses of school effects will be particularly problematic at the high-school level, because (a) the number of high schools in the district is small; (b) assessment data are less complete in high school than in elementary and middle grades; and (c) student populations among the district's high schools vary dramatically (e.g., students entering Westinghouse in ninth grade cannot be easily compared to students entering Allderdice in ninth grade). And information on the characteristics of schools is relatively limited in RTI; further data collection might be warranted to provide a better understanding of how and why some schools are better at producing achievement gains.

6. What characteristics of teachers are associated with gains in student learning? What are the characteristics of teachers who are especially effective at promoting the achievement of low-income and African-American students? Are teachers from particular colleges, or with particular degrees or majors, unusually effective? Are teachers with National Board Certification especially effective? Are particular professional development programs more likely to improve teacher skills in raising student achievement?

The effects of particular teacher characteristics on student achievement might be examined when PeopleSoft data are fully integrated with the RTI system, especially if such data include unique teacher identifiers.

7. How effective are various instructional programs that are operating in the district? For example, what are the effects of the district's efforts to improve achievement in schools that are identified for improvement under NCLB? What are the effects of different supplemental educational services offered to students in the district (whether administered by the district or by an outside provider)?

Data on curriculum implementation (such as those maintained by the Mathematics and Science department, described in detail in Appendix 2) might be integrated with RTI to permit analysis of the relationship between implementation and student outcomes. Assessing the merits of a particular curriculum relative to other curricula, however, requires comparison groups of teachers or schools implementing two different curricula. If PPS chooses to pilot-test alternative curricula, longitudinal modeling methods could be employed to assess the relative achievement benefits of the pilot curriculum and the standard curriculum. To increase the validity of such a comparison, PPS should carefully design the pilot to ensure that an appropriate comparison group can be found within the district.

Also, as noted above, PPS IT staff are in the process of adding codes for various supplemental programs. Once these codes are added to the RTI system, then the impact of these programs can be analyzed—again, with the power of the analysis depending in part on how the users of the programs and comparison groups are selected. This system will accumulate future program participation information for each student. The system is capable of including past program participation information for students, although it may be costly to add this information if it needs to be transcribed from paper records.

8. What are the factors (in terms of behavior, attendance, mobility, achievement, and other variables) that can be measured in elementary and middle grades to identify students who are at high risk of dropping out of school later?

RTI currently includes a substantial amount of information on elementary and middle grade children that can be related to later information on dropping out. Given the uncertainties about dropout data, we believe it might be more appropriate to examine the relationship to ultimate graduation from PPS. However, the ability to answer this type of question is limited by the length of time for which this information has been available. For example, RTI has grades since 1989, which includes the entire grade history (since first grade) for four graduating classes. The grade information indicates which school granted the grades, so we also have information about mobility between school years for this long history. Most information is available for a much shorter time. Assessment, misconduct, citizenship marks, attendance, mobility during the school year, and special education status are available only for the past few years. Of these, the most recent is attendance, which started with the 2003 school year, and the longest is special education status, which goes back to 1998.

9. How do student outcomes differ by neighborhood residence and school assignment patterns?

Using the addresses and feeder pattern information in RTI, longitudinal modeling methods could begin to assess the importance of neighborhood effects. The addition of census data might provide further information about how neighborhood effects operate. Using detailed information on the boundaries of the feeder patterns, it is possible to observe the different outcomes of children who live very close to each other (and therefore probably share many common influences) but attend different schools. Taking advantage of such "natural experiments" provides a valuable opportunity to attempt to distinguish the effect of a school from that of family and neighborhood.

10. How do students with behavior problems affect the learning of other students in their classes?

The combination of misconduct data, classroom grouping information, citizenship grades, and outcome data in RTI should permit an examination of this question. Caution would be merited in interpretation, however, because students are not randomly assigned to classes.

11. Does the district's magnet program provide options that effectively promote student learning?

The use of lotteries to assign spaces in oversubscribed magnet programs in the district creates an opportunity for an especially rigorous and powerful evaluation of magnet program effects, using a randomized experimental design—the “gold standard” in evaluation. Attendance in a magnet program is available back to 1998. There is detailed information in RTI about the magnet assignment process, including information about all applicants, lottery participants and enrollment decisions back to 2002.

12. How does teacher attendance affect achievement? Do the characteristics of substitutes matter? Does the effect of teacher attendance on achievement depend on the pattern of teacher absences? During what time in the semester are absences most damaging to students? Are multiple short absences or one extended absence more damaging?

These questions might be addressed in some depth after PeopleSoft data are integrated with RTI. PeopleSoft has detailed attendance records for teachers, permitting an analysis of whether teacher absences, either intermittent or prolonged, have an impact on student outcomes. The district also maintains a separate Substitute Employee Management System that tracks which substitute teacher is replacing which absent teacher on a daily basis. Using data from this system in conjunction with RTI and PeopleSoft data should permit analysis of the impact of substitute assignment policies on student outcomes, depending on how such assignment policies vary across schools within the district. Some substitute effects, however, will be difficult to measure, because the use of substitutes may vary substantially within a school year, while most of the student achievement measures are annual.

13. How do the various information systems that have been implemented at the district affect student outcomes? Can the improvement of the measurement of outcomes that an information system facilitates be separated from any effect on the outcomes themselves?

One major objective of RTI is to provide useful data directly to principals and teachers. The district has created an interface designed to promote use in the schools, and has initiated professional development for educators on the use of the system. RTI has extensive user logs that track which educators are using it. Therefore, it should be possible to examine whether teachers and principals who use RTI show improvements in their capacity to raise student achievement.

Recommendations for additional data elements for RTI

As the list of questions above indicates, RTI already includes sufficient data to permit a wide range of longitudinal-modeling analyses to be undertaken. Nevertheless, there are a number of additional data elements that could be added to RTI to further increase its utility for informing data-driven decisionmaking. We suggest the following additions to the RTI data system:

- As is currently planned by PPS staff, unique teacher identifier codes should be included in the PeopleSoft data system and imported to RTI, along with the various teacher qualifications and characteristics that are already included in PeopleSoft. Those unique codes should be maintained for each teacher over time. In the current RTI system, teachers cannot be tracked when they move between schools, and teacher codes within schools are re-assigned to new teachers. Unique teacher IDs are needed for a complete and comprehensive analysis of the effect of teachers and teacher characteristics on student outcomes. In addition, unique teacher IDs can permit an assessment of the effects of teacher professional development programs.[6]

- Available data from curriculum content departments should be integrated into RTI. This would include information on participation in professional development as well as information on curriculum implementation, if it is deemed valid and reliable. For example, PPS currently possesses implementation ratings in math and science that might prove useful in assessing the relationship between curriculum implementation and student achievement.
- Data from the existing Substitute Employee Management System should be integrated with RTI, to permit the possibility of examining the effects of different assignment policies for substitute teachers.
- To the extent possible, PPS should consider collecting and including additional baseline data on students entering its schools for the first time, in kindergarten or pre-kindergarten. Such data might include health information, family history and structure, and early academic, social, and psychological assessments.
- To the extent possible, PPS should consider collecting and including additional data on student health status, focusing especially on issues that may affect achievement or attainment (such as, for example, pregnancies or experience of violence).
- If PPS wishes to systematically examine the effect of its technology investments on student achievement, it should collect and include in RTI additional

[6] PeopleSoft currently includes unique teacher identifiers in the form of Social Security numbers. In order to ensure the protection of confidential information, we recommend creating a new set of unique teacher identifiers to be used in RTI.

information about the technology resources available to students, teachers, and schools, as well as information about the actual use of technology.

We should note that the cost of adding more data involves not only collecting and entering the data but modifying the system to incorporate the new data in a useful way. It is possible that new data can leverage existing data both for operational and for policy needs, but to do so requires additional human resources for design and implementation of system modifications. If they make decisions to expand the system, we recommend that the district's leaders weigh the benefits of increased knowledge against both the financial cost of data collection and external support and also the additional burden on the district's Information Technology staff.

Next steps

In sum, the RTI data system currently includes a wealth of information that provides opportunities for rigorous assessments of educational effects, potentially directly informing policy and programmatic decisions in PPS. Select additions to RTI have the potential to enhance this capacity further.

If the district's data are to be used effectively to support educational initiatives, they must be analyzed in ways that are methodologically valid and presented in ways that are meaningful to policymakers and educators. A working partnership between PPS and RAND has the potential to realize the data-driven decisionmaking possibilities inherent in the district's data system. While PPS has a data system with the architecture that is needed to permit rigorous analysis, RAND has research staff with unique skills in longitudinal modeling and VAA. We look forward to discussing with PPS its priorities for analysis, in the hope of establishing a RAND-PPS partnership that could have several major benefits, including:

- Providing critical information for district decisionmaking in support of efforts to raise student achievement across the board and to reduce racial and socioeconomic gaps in achievement;
- Increasing the district's own capacity to analyze and interpret data for instructional improvement;
- Informing major education policy debates of nationwide interest.

Appendix 1: PeopleSoft Data

This overview of the PPS implementation of the PeopleSoft personnel information system is based on discussions with several PPS Information Technology staff who have a working knowledge of the system. Analysis of PeopleSoft data by RAND was outside of the scope of this project, so we have not examined the completeness of the data.

PeopleSoft is the primary personnel data system for PPS. It is not directly linked to RTI, but an initiative is underway to create a unique ID number for each teacher that would permit easy combination of PeopleSoft and RTI data. This initiative is expected to be completed during the summer of 2005.

The PeopleSoft databases contain the following information about teachers, principals and other school-based educational personnel:

Identifying information:

Name	Social Security Number
ID number	Address and Phone number

Demographic information:

Age (Date of Birth)	Race/ethnicity
Sex	

Educational attainment information:

Highest degree	Number of credits toward next degree
Degree granting institution	National Board Certification

Employment information:

Job type (teacher, substitute, counselor, principal, etc.)
Number of years employed in district
Absence record (days of absence each year)
Extended leaves of absence (dates and type of leave for each leave)
Number of years of teaching experience prior to employment in district

Areas of certification
Areas currently teaching
Extra-curricular assignments (dates, type, hours per semester)

Professional development information:
Workshop attendance (type of workshop and dates of attendance IF payroll related)

Substitute teacher assignment information:
Date and school for each assignment

Appendix 2: Math and Science Curriculum Data

This overview of PPS math and science curriculum implementation is based on discussions with several PPS staff from the Math and Science curriculum office. The Math and Science curriculum office does business as PRIME-PLUS (Pittsburgh Reform In Mathematics Education), which is the NSF-sponsored Pittsburgh Urban Systemic Initiative. Analysis of their data by RAND was outside of the scope of this project, so we have not examined the completeness of the data. If reliable data are available and can be linked to students through scheduling information in RTI, such data could contribute to the effort of determining the value added by the various teaching practices.

The district phased in new math curricula beginning in the mid 1990s. Beginning in 1994, first graders used Everyday Math and continued with it throughout their elementary school experience. Beginning in 1996, sixth graders used Connected Math and continued with it throughout their middle school experience. Therefore, by the fall of 1999, every elementary and middle school student was using the new curricula. Because assessment test scores are only available in RTI back to 1999, it is not currently possible to use a pre-post evaluation design on the new curriculum.[7] However, the office reports that the data indicate that there has been uneven implementation of and fidelity to the curriculum. These implementation measures may permit a value-added assessment of the relationship between implementation and achievement, provided the implementation measures are valid and reliable. These data cannot be used to determine how the current curriculum

[7] We understand that the district has student level information on assessment scores is available beginning 1996 but that this information has not been added to RTI. Linking this information to RTI would greatly expand the evaluation options.

compares to a different curriculum. Such a study would require comparable data to be collected while another curriculum was in use.

The science curriculum, Full Option Science System (FOSS), has been used in all elementary classrooms since 2000. Modules of FOSS were phased into middle school classrooms from 2000 through 2002. High schools are at the beginning stages of adopting standards based curriculum.

In this Appendix, we describe the data on implementation and fidelity that have been collected by resource teachers. We have not reviewed the data themselves, only documents associated with data collection. After a description of the data, we appraise their likely merit for contributing to longitudinal models for evaluation.

In addition to the interview, we reviewed the following materials:

- PRIME-PLUS Weekly Log (Resource Teacher activity record)
- Level of Implementation Rating Forms (annual ratings of individual teachers by Resource teachers)
- Instructions to Resource Teachers for filling out Level of Implementation forms
- PRIME-PLUS Annual Report 2003-2004 (including report by an outside evaluator)

Implementation Data

The two main sources of implementation data are the resource teacher (RT) weekly activity logs and the Level of Implementation (LoI) ratings. These are both filled out by the RTs who spend the entire year assisting the teachers in implementing the math and science curriculum. These data sources are available beginning with the 2002-2003 school year.

The weekly logs record the amount of time that the RTs spend with specified ‘audiences’ in each school or in a central location. Each block of time is associated with one of twenty specified activities. The information collected on these logs can be used to determine the amount and type of professional development provided by the RTs to individual classroom teachers. In addition to the activities reported in the RT logs, the Math and Science Office maintains records regarding other teacher professional development activities that are available for analysis.

Although there is some variation between math and science and among the grade levels, the annual LoI form allows the RT to report the following types of information for each teacher:

- Grade taught
- Degree of specialization: either
    - Is teacher a specialist? or
    - Does teacher teach more than two classes in this subject?
- Whether teacher is a long term substitute

- Is subject taught more or less than district recommendation
- Portion of subject time using specified curriculum
- Fidelity to specified curriculum
- Effectiveness in delivering specified curriculum
- Level of confidence in the ratings

The RTs are provided with detailed written instructions on how to complete these forms. The RTs rate every math and science teacher who has been in the classroom at least 75 percent of the year. The final item listed above measures the RT's confidence in the rating. This allows the RT to indicate that some ratings are based on more hours of observation than others.

The PRIME-PLUS annual report provides analysis of the data from both the logs and the LoI forms, including analyses that relate professional development to implementation. There are also analyses of district trends in assessment scores. There is no analysis that relates implementation to scores for specific students, classrooms or teachers.

Likely merit of implementation data for value added assessment

The usefulness of these data for informing longitudinal analyses of achievement effects depends on their validity and on the ease with which they can be linked to RTI.

There are several threats to the validity of the data. The two biggest are data entry errors and the rating process.

Any data are likely to contain some errors that occur in the data entry process. The LoI data are entered directly into spreadsheets by the resource teachers. The use of spreadsheets eliminates the step of transcribing paper and pencil records, which eliminates a possible source of error. However, there are no checks in the spreadsheet to require that only valid values be entered. For example, a resource teacher could accidentally enter "22" instead of "2" for the number of math classes taught. Even though this is not a valid entry, the spreadsheet does not prevent it from being submitted.

This problem is currently addressed through a manual check of the data. The Math and Science Curriculum Office's external evaluation staff look over all the reports from the RTs and all obvious errors are corrected after checking with the RT. Using a database template for data entry could eliminate this step. However, neither manual checking nor database templates can remove errors that do not lie outside the bounds of reasonable answers. Fortunately, some of the data, such as number of math classes taught, can be determined through RTI. This provides a check on the data error rate.

The weekly log data that contains information on teacher professional development is also entered by the RTs into spreadsheets. The biggest potential problem with these data are the use of an unstructured text field ('Audience') to describe the teachers who are involved. This method is prone to produce descriptions of groups that cannot be linked to specific teachers. For example, if the RT writes "some of the math teachers" in this field, this cannot be linked to specific teachers.

In our review of the rating process, we consider possible sources of both bias and imprecision. Measures are biased if they systematically overstate or understate the feature being measured, perhaps for a subset of the population. Imprecise measures may be unbiased in that they are equally as likely to be too high as they are to be too low, but still are of little use if they miss the mark too frequently by too much.

The use of RTs to evaluate classroom teachers can lead to bias if RTs have a conflict of interest. Optimally, a teacher should be rated by observers who have no other relationship to the teacher. In the present case, the RTs work closely with many of the teachers and get to know the teachers and their students over the course of the year. If, for example, RTs give higher curriculum implementation ratings to teachers who they know to be working with an especially difficult classroom of students, then this will bias the measure. Studies have shown that this kind of bias occurs subconsciously and is very difficult to eliminate.[8]

There are several sources of imprecision that are of concern. First, precise measures of implementation require that the assessment protocol be designed to capture the most important aspects of implementation and that the RTs follow the protocol closely. It is beyond the scope of this project for us to examine these issues.

[8] Babcock L, Loewenstein G, Issacharoff S & Camerer C. Biased Judgments of Fairness in Bargaining. The *American Economic Review*, Vol. 85, No. 5. (Dec., 1995), pp. 1337-1343.

Second, the RTs observe some teachers for very little time during the year. The measures of implementation and fidelity are likely to be statistically "noisy" for teachers with whom their RTs are not familiar.

Third, variation among RTs in how they apply the rating scales can lead to noisy measures. PRIME-PLUS has attempted to minimize this source of imprecision by providing written rating instructions to RTs, but without having some teachers be rated by multiple RTs, it is not possible to tell whether the directions are successful in eliminating inter-rater variation.

There is a clear trade-off between bias and precision that arises from using the RTs to measure the implementation of the curriculum. The RTs have the opportunity to observe many of the teachers many times throughout the year. Therefore using the information that RTs acquire through continued observation instead of infrequent visits by independent observers increases the precision of the measure, but at the expense of possible bias discussed above. The only alternative that would be better in terms of both bias and precision would be frequent visits by independent observers, but this alternative is very costly.

While it is important to recognize these weaknesses of these data, we believe that the RT activity logs and LoI ratings have the potential to contribute to longitudinal analysis of student outcomes. Because longitudinal analyses focus on changes in student performance from year to year, some of the biases discussed above will be eliminated.

Imprecision remains a concern, but using confidence ratings and accounting for the assignment of teachers to each RT can minimize the impact of this source of imprecision on the analysis. Only a more thorough investigation can determine whether the remaining bias and imprecision in the measures will prevent a useful analysis.

Another possible concern is related to how implementation data will be used in analyses, rather than the quality of the data itself. Consider, for example, a problem that would arise if teachers who implement the curriculum more completely also have higher-quality teaching methods in other ways. Perhaps they are the teachers who put more energy into all of their teaching activities. In a cross-sectional analysis, better outcomes for the students of these teachers would be attributed to their good curriculum implementation although some, and maybe all, of their students' high performance would really be due to the quality of their general teaching practices and not to the curriculum. This bias in the analysis would exist even if the measure of implementation was unbiased.

Fortunately, longitudinal analysis can address this problem, provided the level of implementation changes over time for the teachers. By comparing the value added by teachers before they learn to implement the curriculum to the value added by the same teachers after they learn to implement the curriculum, we avoid the problem of comparing teachers with generally high quality teaching practices to those with low quality teaching practices.

Finally, we need to determine if the data can be easily linked with RTI data. Only by linking the implementation data with student data can we learn whether proper implementation of the curriculum leads to better student outcomes such as higher assessment scores, higher graduation rates and fewer misconduct episodes. Once the data are linked, we can analyze whether the curriculum is equally effective for all students or whether some groups of students require additional resources.

Both the RT weekly activity logs and the LoI ratings are in electronic form, which would facilitate the linkage process. The LoI ratings will need to be matched to RTI by school and teacher name. Matching on teacher names can lead to some errors, but the current lack of a teacher ID number in the RTI system makes this step necessary. Linking the weekly activity logs will be much more difficult. The column labeled "Audience" presumably contains descriptive information about groups of teachers. We would need to look at a sample of completed logs to determine the ease with which these text fields could be used to identify individual teachers, but we expect that such a linkage would prove to be costly.

Other Possible Uses of the Implementation Data

If shown to be reasonably reliable and valid, the data from the RT activity logs and LoI ratings can be used to better understand both what RTs can do to more effectively promote implementation and whether particular groups of teachers (i.e. particular schools, levels of seniority, degrees of specialization, etc.) require extra attention.

Appendix 3: SchoolNet Data

This overview of PPS implementation of the SchoolNet system is based on discussions with several PPS staff. Analysis of their data by RAND was outside of the scope of this project.

The purpose of this overview is to determine whether the SchoolNet system has currently or could in the future have useful data on variation in teaching practices within the district. If reliable data are available and can be linked to students through scheduling information in RTI, such data would contribute to the effort of determining the value added by the various teaching practices.

SchoolNet is an instructional management system for teacher decision-making. The core of SchoolNet consists of three modules:

- Account: contains student information system data that is pushed in from RTI
- Align: contains information on curriculum and standards information that is loaded by PPS staff
- Outreach: communications module that allows administrators and teachers to create web pages, bulletin boards and calendars for their school or their classes, usable by students and/or parents

Implementation

- Account

The data bridge has been built that sends many student records to SchoolNet from RTI. It sends attendance, schedules, grades, marks and non-local assessments. It currently does not send misconduct information, magnet information or local assessments.

- Align

Curriculum information has been loaded for benchmark years (grades 3, 5, 8 and 11) for math, reading and science. This required that the curriculum be broken into modules and aligned with standards. Each small unit of the curriculum has been tagged to associate it with the standards that it addresses. This was finished by December 2004.

During the spring of 2005, one person from each building was given intense training and became certified on SchoolNet. This person will assist with training the rest of the building personnel. In the fall of 2005, principal teams from each building (approx. 2 staff including the principal) will be trained with an emphasis on generating reports.

The Math Department made a special request. They wanted all teachers to be able to access curriculum during the spring of 2005. IT set up a 'generic' account that permits all teachers to access on-line curriculum but does not link to student data nor keep track of lessons taught or standards covered. Also, Math opted for loading grades 1-5 curriculum rather than benchmark years.

Notable features:

- In SchoolNet, school administrators and teachers can create their own reports, read pre-formatted reports as well as access reports designed by central administration. RTI centralizes the report design function. In RTI, teachers and administrators can read pre-designed reports but cannot design their own.

- Teachers log in and see the curriculum for the classes they are teaching. They see information about what standards go with what lesson units. They can ‘drag and drop’ lessons and see how they are meeting their standard targets
- For each curriculum unit, they have on-line access to lesson outlines. Teachers can drill down into an outline for the entire year to get to unit and lesson info including resources such as worksheets, video clips, and assessments.
- A log is kept for teachers and administrators that shows how much of each standard is taught by each teacher in each class. (This changes the status quo in which a written lesson plan is given by the teacher to the principal each week and no record is kept of whether particular standards are covered.)

If the SchoolNet system is fully implemented for the district, it would provide an opportunity to evaluate the relationship between the use of particular teaching activities and student achievement. By linking teacher logs to student data in RTI, it would be possible to determine whether spending more time teaching particular standards is associated with higher scores on assessment items related to those standards. The validity of such an analysis would depend critically on the accuracy and validity of the information recorded in the teacher logs; assessing those logs was beyond the scope of this study.

SchoolNet may also provide an opportunity to investigate whether the use of information technology improves outcomes. If SchoolNet is fully implemented, it may be possible to evaluate whether students whose teachers are using SchoolNet's functions more fully have assessment scores that increase more than other students. This suggests that a cost benefit analysis of the technology may be possible.

Appendix 4: Assessment Data Available in RTI

Table 1 reports the year by grade availability of student level standardized test data in RTI for all subjects. By tracing a diagonal sequence of this table, it is possible to see all the tests that are available for a particular set of students making normal progress.

Table 1: Student-Level Annual Assessment Results in RTI

Grade	1999	2000	2001	2002	2003
1			BAM		TN_Lang, TN_Read, TN_Math
2			BAM, SAT9Mult, SAT9Open		TN_Lang, TN_Read, TN_Math
3			BAM, SAT9Mult, SAT9Open		
4	NSELA, NSMath	NSELA, NSMath	NSELA, NSMath, SAT9Mult	NSMath	NSMath, TN_Lang, TN_Read, TN_Math
5	PSSA_Read, PSSA_Math	PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math, SAT9Mult	PASS_Sci, PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math
6			BAM, SAT9Mult, SAT9Open	TN_Math	PSSA_Write, TN_Lang, TN_Read, TN_Math
7	NSELA, NSMath	NSELA	NSELA, BAM	TN_Math	TN_Lang, TN_Read, TN_Math
8	PSSA_Read, PSSA_Math	PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math
9				TN_Math	PSSA_Write, TN_Lang, TN_Read, TN_Math
10	NSELA	NSELA, NSMath	NSELA, NSMath	NSELA, NSMath	NSELA, NSMath
11	PSSA_Read, PSSA_Math	PSSA_Read, PSSA_Math	PASS_Sci, PSSA_Read, PSSA_Math, PSSA_Write	PASS_Sci, PSSA_Read, PSSA_Math, PSSA_Write	PASS_Sci, PSSA_Read, PSSA_Math, PSSA_Write

Notes: Year is school year: e.g. 1999 represents school year 1999-2000.
Following a diagonal sequence provides information on students who are making normal progression. Two such sequences have been highlighted in purple.

Due the limited amount of time, we restricted our analysis of data completeness to math assessment scores. RTI contains student-level score information for three math assessment tests: PSSA, New Standards and Terra Nova. The data system also contains item level information for each student for BAM, but does not aggregate these items into a score for each student. RTI also includes many reports that give aggregated score statistics. For example, there are reports that give statistics for each school, for racial categories, and for the district as a whole.

Table 2 provides information about the completeness of the data for math assessment tests. Data availability generally improves in recent years, and the data are the most complete for PSSA. Over 90% of fifth and eight grade students have score information for 1999 through 2003. PSSA reporting rates are somewhat lower for eleventh grade students, with 2000 being substantially below the other years. New Standards scores are available for over 80% of fourth grade students from 1999 to 2003, with 96% available in 2003. Tenth grade New Standards availability has increased from 34% in 2000 to 77% in 2003. Terra Nova scores are available for over 95% of first, second, fourth, sixth and seventh grade students in 2003. Lower reporting rates prevail for ninth grade students and for fourth, sixth, and seventh grade students in 2002.

The reduced completeness in higher grades may be the result of at least two factors. First, absenteeism is higher in upper grades, which may lead more students to miss testing dates. Second, these data include students who drop out during the course of the school year but prior to testing dates. In consequence, these figures may be somewhat conservative (understated) and should *not* be interpreted as accurate estimates of test participation for purposes of compliance with state and federal requirements. Nevertheless, they are useful guides for assessing the extent to which longitudinal growth modeling will be possible in PPS.

Table 2
Math assessment test data availability
Percent of students for whom a score is recorded
(All years are fall semesters, e.g. 1999 is 1999-2000 school year)

PSSA

Year	Grade 5	8	11
1999	94.0%	93.4%	75.0%
2000	97.0%	92.3%	54.2%
2001	96.8%	91.6%	88.7%
2002	95.8%	90.5%	82.9%
2003	95.2%	94.2%	84.0%

New Standards

Year	Grade 4	7	10
1999	83.7%	86.2%	
2000	80.6%		33.6%
2001	79.5%		54.5%
2002	83.5%		72.6%
2003	96.4%		76.7%

Terra Nova

Year	Grade 1	2	4	6	7	9
2002				81.2%	81.3%	67.4%
2003	96.9%	96.6%	96.0%	97.7%	97.5%	74.3%