Examining Relationships Among Assessment Scores and Math Coursework in an Urban School District

Carol S. Parke

Duquesne University

Abstract

This study investigates relationships between assessment scores and other indicators of math performance. The impetus for the research came from a district's need to better understand high school math achievement. Longitudinal data for a cohort of students were obtained, including math scores from their state assessment, TerraNova, and New Standards Reference Examination; cumulative math GPA; number of math courses taken; and type of math courses taken. The paper illustrates how researchers can help districts utilize their extensive databases to proactively examine data beyond accountability requirements. A discussion focuses on how results helped target areas for improvement and identified further analysis within schools.

Introduction

In this era of accountability, school districts are required to maintain extensive longitudinal student databases complete with information including attendance, demographics, mobility, discipline, state test scores, course enrollment, and grades received in courses. Data systems created by districts are only useful in transforming schooling when they provide meaningful data that stakeholders can use to raise questions, identify issues, and make informed decisions (Schmoker, 2008).

The research described in this paper stems from a partnership between a large urban school district, a community educational organization, and a local university. The partnership's initial focus was to create annual School Progress Reports (A+ Schools, 2007) that allowed administrators, teachers, and parents to access a variety of demographic, contextual, and performance indicators for each school in a form that was not available elsewhere. The data served as starting points for discussion about the strengths of each school as well as the challenges faced. Supplementary analyses followed the release of each Report with the purpose of further examining specific areas of interest to the district, such as attendance and mobility (Parke, 2006; Parke, 2008; Parke, 2009; Parke and Kanyongo, in press). The study described here is from an analysis undertaken due to the growing concern regarding low math scores on the state's 11th grade assessment. A broad question raised by the district and community was "how do students' math scores on the Pennsylvania System of School Assessment (PSSA) relate to other available measures of mathematics performance?" To this end, the analyses examined relationships between student achievement on the state test and five additional math indicators.

Although the study focuses on one school district, the purpose for sharing this research with the assessment community is to provide an example of how researchers can help districts better utilize their extensive databases to explore questions of interest to them, highlight areas that need attention, and make proactive decisions to ultimately improve learning for all students. The capacity of student data to make improvements in districts is quite large; unfortunately, much of it remains untapped because of a lack of time in personnel's busy work days, a lack of resources, or a lack of knowledge.

The following review of literature begins with a description of the state's research on relationships between assessment scores and course grades and is followed by correlation studies between scores and grades at the national level. Then, research from the mathematics education field is described in terms of the importance of incorporating coursework variables in studies that examine math achievement. Finally, the influence of race and gender on mathematics achievement is discussed.

Literature on Test Scores, Courses, and Grades

Previous Research on the PSSA

The PSSA is a standards-based, criterion-referenced assessment that measures student outcomes according to state standards. It consists of both multiple-choice and open-ended items, and during the years of this study was administered in grades 3, 5, 8, and 11. Evidence for reliability, validity, and item evaluation is available in yearly technical manuals (e.g., Pennsylvania Department of Education, PDE, 2005.) In all these respects, the PSSA for mathematics is shown to be a technically sound assessment.

Studies on the 11^{th} grade assessment investigated relationships between PSSA scores, SAT scores, self-reported total grade point average (GPA), and math course grades (Koger, Thacker, & Dickinson, 2004). Convergent validity coefficients between the PSSA and SAT were high (approximately .850.) Although the two assessments differ in content, format, and purpose, students who did well on the PSSA tended to do well on the SAT. Relationships between the PSSA and the two self-reported measures of grades were also significant but lower in magnitude (.546 for GPA total and .534 for math course grades.)

When studying PSSA scores by demographic subgroups, Koger et. al. (2004) found that, on average, high school males performed significantly higher than females, with an effect size of d = .31. White students performed higher than Black students (d = 1.07). Students not from economically disadvantaged households performed higher than students from disadvantaged households (d = .75). Comparisons for total GPA were significant and in the same direction as the results for the PSSA, but with smaller effect sizes for ethnicity (.71) and economically disadvantaged students (.39). Gender results were in the opposite direction. Mean GPA total was significantly higher for females than males (d = .239).

Correlations Between Test Scores and Grades

Over the past few years, Zwick and colleagues (e.g., Zwick & Green, 2007; Zwick & Schlemer, 2004; Zwick & Sklar, 2005) conducted numerous studies on the SAT and grade point averages in high school (HSGPA) and the first year of college (FGPA) to determine if relationships among scores and grades were consistent across demographic subgroups. If the relationship is stronger in one student subgroup compared to another, then the prediction of test scores using the grade variable is less effective for the subgroup with the weaker correlation. Zwick and Schlemer's study in 2004 focused on the effectiveness of SAT scores and HSGPA to predict FGPA. Using a single regression equation for the entire cohort, average prediction errors for each subgroup were obtained. Substantial overpredictions occurred for Latino non-native speakers when high school grades were the only predictor of college performance. After incorporating SAT into the model, prediction errors were smaller. A notable degree of overprediction also occurred for Asian Bilingual, Filipino, African-American, and Latino/English groups. Underpredictions were more common in the White group. Actual FGPA was higher than what was predicted by SAT and HSGA.

Zwick & Schlemer (2004) also estimated separate regression equations for each subgroup. The total amount of explained variance in SAT using all three predictors (HSGPA, SAT math, and SAT verbal) was somewhat small, ranging from .15 to .25 for most groups, with the exception of .44 for the Asian/English group. Similarly, Zwick & Sklar (2005) showed that about 23% of the variance in FGPA was explained by the high school grades and SAT.

When estimating correlations, two methods may be used. The most common is to combine data from all students, without considering the school attended, and obtain the across-school correlation matrix. This matrix represents within-school and between-school associations between variables. An alternative method is to obtain pooled, within-school correlations. This matrix does not reflect between-school variations. Using data from the College Board, Zwick and Green (2007) investigated the two methods. Across-school matrix results indicated that relationships between SES and SAT were substantially higher than relationships between SES and HSGPA. After removing between-school variations, the within-school matrix showed that the SES and SAT relationships were similar to the SES and HSGPA relationships. However, there was not a difference in the two methods for correlations between SAT and HSGPA (within-school correlation was .525 and across-school correlation was .513.)

Willingham, Pollack, & Lewis (2002) sought to understand why the relationship between test scores and grades is only moderate at best. A potential reason for the moderate relationship is the inherent nature of each measure. In terms of content and statistical properties, a standardized assessment is developed to provide data comparable across schools. Course grades, on the other had, can vary widely across schools and teachers not only because of variations in content and format but also because teachers may take into account elements beyond knowledge and skills (e.g., class participation, attendance, behavior, and effort). In their analysis of data from NELS

1992 transcript files, three major factors accounted for differences in observed grades and grades predicted from test scores: 1) grading variation among schools, 2) scholastic engagement (e.g., showing initiative in school), and 3) teacher ratings (influence of additional elements in evaluating achievement.)

Incorporating Coursework Variables

Studies in educational measurement (e.g., ACT, 2004; Campbell, Hombo, & Mazzeeo, 2000; CEEB, 2001) and mathematics education (e.g., Ma, 2000; Ma & Wilkins, 2007; Riegle-Crumb, 2006; Wilkins & Ma, 2002; Wilkins, Zembylas, & Travers, 2002) incorporated coursework indicators into studies on academic achievement. Within the measurement field, results from NAEP trends analysis (Campbell, Hombo, & Mazzeo, 2000) and profiles of college-bound seniors from the College Board (CEEB, 2001) show strong relationships between math courses students take in high school and their achievement. Moreover, reports published by ACT on relationships between high school math coursework and future success in college show that "not only is taking the *right number* of courses important, but taking the *right kind* of courses is critical to student readiness for college-level work" (ACT, 2004, p. v).

In mathematics education, a body of research by Ma and Wilkins (Ma, 2000; Ma & Wilkins, 2007; Wilkins & Ma, 2002) focused on the influence of coursework on achievement in middle and high school. Using cohort data from 7^{th} to 12^{th} grades, Ma (2000) investigated the impact of taking specific math courses, such as prealgebra, geometry, and calculus, on students' attitudes and achievement in math. After accounting for prior achievement, attitude, gender and SES, results from regression analysis indicated that taking algebra 1 in Grade 11 (considered to be a low-level course at this grade level) did not have a significant impact on achievement. However, taking algebra 1 in Grade 8 and trigonometry in Grade 11 (both courses considered advanced for the particular grade levels) showed substantial effects. Thus, the timing of math courses appears to impact achievement. Smith (1996) also found that students who take algebra prior to high school had higher 10th grade math scores than students who took algebra during high school.

In their growth study, Wilkins and Ma (2002) incorporated several student personal factors into the model, such as math self-concept, educational aspirations, home resources, peer influence, teacher/parent encouragement, exposure to books, and time spent on homework. Student factors related to growth differed in middle versus high school. Self-concept had a strong effect on middle school growth, whereas educational aspirations effected high school growth. Peer influence was related to growth in middle school, but not in high school.

Finally, Ma and Wilkins (2007) investigated the extent to which math coursework influences growth in math achievement from 7th to 12th grades. In general, low-level courses had the smallest impact on growth, and advanced courses had the largest impact. Coursework effects did not systematically bias demographic subgroups. Success (not failure) in prealgebra and algebra courses in middle school was important in maintaining future growth in achievement.

Race, Gender, and Mathematics Achievement

There is a wealth of research on the varying math achievement levels among demographic subgroups of students. According to Campbell, Hombo, and Mazzeo (2000), trends in math achievement on the National Assessment of Educational Progress (NAEP) over the past three decades show that the average math score of White students is higher compared to their African American and Hispanic peers. Overall, the gap decreased between 1973 and 1999, but a significant difference remains. Research also shows that math scores of students from both ethnicities varies by family income status. In a study of urban middle schools, Kinney (2008) found that 4th, 6th, and 8th grade students who qualified for free/reduced lunch, a proxy for socioeconomic status, had significantly lower math achievement than those who did not qualify. With regard to gender, however, research is somewhat inconclusive. Some investigations report differences between males and females, while others do not. Examples from a few specific studies on the influence of demographics on math achievement are described below.

A state-level study on gender differences at the high school level (Koger et al, 2004) is one example of mixed results. Math achievement as measured by the state assessment showed that males had significantly higher scores than females. Conversely, when math achievement was measured by course grades, females came out on top with higher grades than males. An international study of high school students across 16 countries (Wilkins, Zembylas, & Travers, 2002) examined whether differences in math literacy were due to school variables (such as opportunity and experience) or individual student characteristics. Both gender and self-concept were found to be two of the most important predictors of math success. Males had higher scores than females, and higher math self-concept was related to higher math scores.

The release of the *Curriculum and Evaluation Standards for School Mathematics*, published by the National Council of Teachers of Mathematics (1989), spurred much research in examining math performance by students' race/ethnicity. The Standards stated that all students can learn mathematics and called for an increased emphasis on mathematical communication, problem solving, reasoning, and connections. A four-year study of elementary students conducted by Pungello, Kupersmidt, Burchinal, & Patterson (1996) examined ethnicity, gender, and socioeconomic status. Math achievement was negatively associated with the minority student group, specifically African American. When analyzing interaction effects, Black students had a smaller gap between the two income groups compared to White students. In another study, the conceptual and computational scales from the California Achievement Test (CAT) were used to measure math achievement (Hall, Davis, Bolen, & Chia,1999). No gender differences were found for either subscale. However, White students had higher scores than Black students, especially on the math concepts scale. They also found that parent variables, such as educational level and math anxiety, were related to math scores and varied somewhat by race.

Within the past decade, indicators of students' math achievement expanded from using only test scores to incorporating information about math courses taken and grades received. Several studies found differences in the types of courses taken across ethnicity subgroups. In a study by Byrnes (2003), White students were more likely to take classes beyond algebra (such as geometry and trigonometry) when compared to their African American or Hispanic peers. Riegle-Crumb (2006) investigated high school math course patterns by gender and race. White students of both genders had higher representation in advanced courses when compared to African American and Latino students of the same gender. In addition to taking fewer classes, these two student subgroups had higher failure rates when compared to White and Asian peers of the same gender. Furthermore, African American and Latino students of both genders had smaller percentages of students obtaining high grades in their math courses.

Summary of Literature

The research on test scores, courses, and grades show that the type of math course taken is more important than the quantity of courses taken when examining students' readiness for college(ACT, 2004). In general, a student will more likely be an achiever (i.e., have high test scores) if he/she takes algebra in middle school with a positive self-concept, then goes on to take advanced courses in high school and has aspirations to attend college (Wilkins & Ma, 2002). Literature also shows that high-achieving students, those who perform well on state mathematics tests, tend to perform well on college-readiness tests (Koger, Thacker, & Dickinson, 2004) and are better prepared for future success in college (ACT, 2004). Moreover, research shows the importance of examining data by demographic subgroup to identify potential inequities in assessment (Zwick & Sklar, 2005) and the availability of advanced courses (Wilkins & Ma, 2002). An implication of these research results is that when students do not have access to high-quality, advanced math courses, their achievement and options for future careers become limited (Ma & Wilkins, 2007).

Purpose and Research Questions

Due to the heavy emphasis on accountability and the need to document Adequate Yearly Progress, all states and, increasingly, some districts are maintaining a wealth of student information in electronic databases. These longitudinal systems contain data to "determine not just whether an individual student's performance is improving, but also how and why." (Data Quality Campaign (DQC), 2009). The goal of this analysis was to investigate the nature of relationships between a state assessment and other indicators of math performance in order to provide an urban school district with a broader picture of students performance than the data they use to meet accountability requirements mandated by the No Child Left Behind Act.

This paper is unique in that the impetus for undertaking the research came from district concerns about low math performance on the state test. Year after year the district received results identifying gaps in demographic subgroups. Longitudinal student information regarding course-taking and math grades had not been systematically examined. Thus, the following three questions were posed. Personnel were specifically interested in knowing more about mathematics performance for students who stayed in the district's high schools, therefore analyses were conducted on data from a cohort of students who attended the district's high schools from 9th grade in 2002-03 to 11th grade in 2004-05.

1) What are the relationships among scaled scores on the TerraNova (TN) in 9th grade (2002-03), New Standards Reference Examination (NS) in 10th grade (2003-04), PSSA Math in 11th grade (2004-05), cumulative grade point average for math courses (GPA Math), number of math courses (Course Total), and type of math courses (Course Type)?
2) Do the relationships above remain consistent across gender, ethnicity, and socioeconomic (SES) subgroups?
3) What proportion of variance in 11th grade PSSA math scores is explained by 9th grade TN scores, 10th grade NS scores, GPA math, Course Total, and Course Type? And, are the results similar across ethnicity subgroups?

Methodology

Sample

This urban school district serves the second largest city in one northeastern state. Total student enrollment was approximately 32,000 during the time of the study. Across all grade levels, the majority of students (57%) were African-American, 38% were Caucasian, and 6% were Asian, Hispanic, or American Indian. Two-thirds of all students (64%) were eligible for free/reduced lunch.

There were 53 elementary schools in the district. A small portion of these schools also served grades 6, 7, and 8. Average student enrollment was 287. The 17 middle schools, grades 6 through 8, had an average enrollment of 383. Average enrollment in the 10 high schools, grades 9 through 12, was 981. Similar to most school districts in urban areas, student mobility was high as well as the number of student disciplinary infractions, especially in the upper grade levels.

Specifically in the high schools, average student attendance was somewhat low (82%). Scores on the state assessment at grade 11 were below the state average (14 percentage points for reading and 13 percentage points for mathematics). A disparity also existed between scores for Black and White students. Across all high schools in the district, 59% of White students scored at the proficient or advanced level, whereas only 17% of Black students scored at these two highest levels. The state also had a disparity between the two subgroups, although not as large.

The cohort of district students focused upon in this paper is defined as all students who attended the district's ten high schools as 9th graders in 2002-03, 10th graders in 2003-04, and 11th graders in 2004-05 and who took the three large-scale math assessments. This represents a total of 1,298 students. Approximately 42% of the cohort students were Black, 55% were White, and 3% were other ethnicities (Asian, Hispanic, or American Indian). Slightly less than half the students (43%) were eligible for free or reduced lunch. The remainder of students in the district across these school years were called the "non-cohort" and were not included in the analysis described here.

When compared demographically to non-cohort students (Parke & Keener, 2011), the cohort had significantly higher percentages of female students, White students, and students not from low-income families as compared to the non-cohort. Academically, the cohort had significantly higher mean scores on the large-scale assessments at each grade level than the non-cohort.

Data Source and Variables

Data for the study was obtained from the district's Real-Time Information system, a web-based interface designed to provide efficient and accurate access to the school's server. The district appears to be ahead of other districts across the country in terms of the potential of its system to provide meaningful longitudinal data to stakeholders (Brooks-Young, 2003; DQC, 2009; Enomoto & Conley, 2007). Several features make it a strong database. First, all information is consolidated into a centralized location. Many school systems keep records in multiple locations, leading to inaccurate data. Secondly, one department in the central office is responsible for developing and maintaining the database, and it is staffed with people who have assessment, data management, and computer/technical experience. Third, training and support for teachers and clerical staff in using the database is offered on a regular basis.

Variables in this study include math scaled scores on three large-scale assessments: the TN in 9^{th} grade, NS in 10^{th} grade, and PSSA in 11^{th} grade. The district had been administering the TN and NS in order to have standardized information about students' math performance prior to the state test. There are also three math coursework variables: 1) cumulative GPA for math courses, 2) total number of math courses taken , and 3) type of math courses taken. The type of math course was a dichotomous variable: core courses only versus core plus advanced courses. Core courses included algebra 1, algebra 2, and geometry. Advanced math courses included elementary functions, advanced topics, linear algebra, calculus, and statistics. Demographic variables include gender, ethnicity, and eligibility for free/reduced lunch (a proxy for SES.)

Data Analysis

Correlation analyses was used to answer the first question regarding the relationships among achievement indicators for the entire cohort. To investigate the second question, correlation coefficients were obtained separately by gender (male and female students) and by ethnicity/SES subgroups (Black free/reduced lunch, Black regular lunch, White free/reduced lunch, and White regular lunch students). Fisher's r-to-z transformation was used to determine if correlations between subgroups were significantly different. The "other" ethnicity subgroup was too small to include in the analyses.

Multiple regression analyses were used to answer the third question regarding the amount of variance in PSSA math performance explained by other mathematics indicators. The first analysis entered demographic variables in Step 1 and the five math indicators in Step 2. The second analysis examined the unique information provided by the two sets of math indicators Assessment variables and math coursework indicators were entered into the equation in different orders. In other words, one regression added the TN and NS in Step 2 of the model and GPA Math, Course Total, and Course Type in Step 3. The other regression reversed the order by adding the three coursework variables in Step 2 and the two assessments in Step 3. Changes in R^2 were important to examine because the district was interested in knowing how much of the variance in PSSA scores could be explained by coursework information without knowing students' scores on the other assessments. The final analyses examined whether the strength of the prediction differed by ethnicity subgroups. For example, do coursework variables account for a larger amount of variation over and above TN and NS for one ethnicity subgroup compared to another? Thus, separate regression equations were estimated for Black students and White students.

These data analyses techniques were chosen over other equally appropriate procedures because the purpose of conducting this research was to help the district better understand mathematics achievement and produce results that were meaningful to them. A final note is that when students are nested in schools, traditional regression procedures involving ordinary least squares analysis may be problematic because of the assumption of independence of observations (Goldschmidt, Martinez, Niemi, & Baker, 2007). If this assumption is not tenable, then hierarchical linear modeling is the desired statistical procedure. Intraclass correlations can be used to examine this assumption by determining if variances in the outcome variable attributed to schools is large. When these correlations are large, then traditional regression has a tendency to underestimate standard errors. In this study, intraclass correlations, regardless of whether they were obtained by a random effects or mixed model, were less than .01, a level which is considered to satisfy the independence assumptions. Therefore, the traditional regression procedures described in the above paragraphs were deemed appropriate for the data in this study.

Results

Research Question 1: Relationships Among Indicators for the Entire Cohort

Correlations among all math indicators are shown in Table 1. TN and NS are strongly related to PSSA scores. The largest correlation occurred between the NS and PSSA ($r = .859$). GPA math and Course Type were also significantly correlated to PSSA, and the magnitude of the coefficients were moderately large ($r = .672$ and $r = .557$, respectively). Course Total was not significantly related to PSSA (r = -.024). Intercorrelations among math indicators were moderate to strong with the exception of course total.

Table 1. Correlation matrix for scores on three mathematics assessments and three math coursework indicators.

	PSSA	**TN**	**NS**	**GPA Math**	**Course Total**	**Course Type**
PSSA	---					
TN	.780*	---				
NS	.859*	.798*	---			
GPA Math	.672*	.524*	.662*	---		
Course Total	-.024	-.032	-.057	-.111*	---	
Course Type	.557*	.522*	.612*	.403*	.102*	---

* p<.001

It is possible that the correlations are underestimates of the true relationships because they were calculated at the student level. Differences between schools in assessment performance and in the assignment of course grades tends to lower the correlations (Willingham et al, 2002). Therefore, within-school correlations were also computed. The pooled within-school correlations were quite similar to the across-school correlations given in Table1. Over 80% of the differences were less than .030, and many were .010 or less. For example, the across-school correlation between the PSSA and GPA Math was r = .672, and the within-school correlation was r = .671. One reason for the similarity in relationships estimated by the two methods might be that the data was from one large school district, whereas other research studies have used national data from many school districts across states. Most likely, the degree of grading variation and course-taking patterns between schools would be higher at the national level than at the district level.

Research Question 2: Relationships Among Indicators by Subgroups

To examine whether the magnitude of relationships was consistent across demographic subgroups, separate correlations were obtained for gender (male, female) and four ethnicity/SES categories (Black free-reduced, Black regular, White free-reduced, and White regular lunch).

The correlation matrix for gender is not shown here since there were negligible differences in male and female correlations among all indicators. Relationships between PSSA and other math indicators were equally strong for both genders, with the exception of PSSA and Course Total which was equally weak for both genders. When comparing coefficients for free/reduced and regular lunch students within ethnicity subgroups, Table 2 shows similarities and differences. First, relationships between PSSA and other indicators (except course total) were stronger for regular lunch than free/reduced lunch students, regardless of ethnicity.

For example, the PSSA and TN correlation was .755 for White regular lunch and .711 for Black regular lunch students, whereas the correlation was .661 for White free/reduced lunch and .645 for Black free/reduced lunch students. Using Fisher's r to z transformation, the correlation between PSSA and TN among all White regular lunch students ($r=.755$) was not significantly different from the correlation between PSSA and TN among all Black regular lunch students ($r=.711$), $z = -1.10$, $p>.05$. Likewise the PSSA/TN correlation among all White free/reduced students ($r=.661$) was not significantly different from the PSSA/TN correlation among all Black free/reduced students ($r=.645$), $z = .30$, $p>.05$. Similar results were found for the PSSA relationships with three of the four remaining math indicators (NS, GPA Math, and Course Total).

However, the relationships between PSSA and Course Type did differ for the two ethnicities. The PSSA/Course Type correlation among all White regular lunch students ($r=.587$) was significantly different from the PSSA/Course Type correlation among all Black regular lunch students ($r=.412$), $z = -2.73$, $p<.01$.

Table 2. Correlation matrix for scores on three assessments and three math coursework indicators by ethnicity and SES.

	PSSA	TN	NS	GPA Math	Course Total	Course Type
PSSA	---					
TN						
Black, free/red[1]	.645*	---				
Black, regular	.711*					
White, free/red	.661*					
White, regular	.755*					
NS						
Black, free/red	.731*	.664*	---			
Black, regular	.794*	.766*				
White, free/red	.806*	.644*				
White, regular	.829*	.774*				
GPA Math						
Black, free/red	.579*	.347*	.543*	---		
Black, regular	.617*	.521*	.588*			
White, free/red	.513*	.266*	.517*			
White, regular	.633*	.489*	.631*			
Course Total						
Black, free/red	-.001	-.029	-.107	-.174*	---	
Black, regular	.049	.052	.024	-.041		
White, free/red	-.160	-.061	-.121	-.288*		
White, regular	.039	.032	.018	.013		
Course Type						
Black, free/red	.298*	.225*	.341*	.263*	.016	---
Black, regular	.412*	.429*	.463*	.332*	.229*	
White, free/red	.400*	.331*	.493*	.121	.072	
White, regular	.587*	.579*	.664*	.401*	.168*	

* p<.001

[1]Sample sizes are 364, 182, 166, and 545 for Black, free/red; Black, regular; White, free/red; and White, regular lunch students, respectively.

Next, SES coefficients were compared within each ethnicity subgroup. For the Black subgroup, regular lunch correlations were higher than free/reduced lunch correlations for each PSSA relationship, but differences were not statistically significant.

Within the White subgroup, regular lunch correlations were higher than free/reduced lunch correlations for all PSSA relationships with other indicators, and they were also statistically significant for four of the five pairs of correlations (PSSA/TN, PSSA/GPA math, PSSA/Course Total, and PSSA/Course Type). For instance, the correlation between PSSA and Course Type among all White regular lunch students (r = .587) was significantly different from the correlation between PSSA and Course Type among all White free/reduced lunch students (.400), z = 2.80, p<.01.

In summary, there were no statistically significant differences in the strength of PSSA relationships with other math achievement indicators for the Black regular lunch group versus the White regular lunch group, except for course type which had a stronger relationship for the White subgroup. When comparing SES categories, regular lunch correlations were always higher than the free/reduced lunch correlations. Most of the differences in these correlations were statistically significant within the White student subgroup but not the Black student subgroup.

Research Question 3: Explaining Variance in PSSA Math Performance

Multiple regression analysis was used to answer the third research question regarding the amount of variance in 11th grade PSSA math scaled scores explained by demographic variables, 9th grade TN and 10th grade NS math scaled scores, GPA Math, Course Total, and Course Type.

The first model included only the demographic variables. Ethnicity, SES, and gender accounted for 26.5% of the variance in PSSA scores. Ethnicity and SES were significant predictors (p<.001), but gender was not. In the second model, ethnicity and SES were entered as a block in Step 1, then all five math indicators were entered in Step 2. The indicators accounted for an additional 52.1% of the variance above and beyond the demographic variables. Thus, the full model explained a total of 78.5% of variance in PSSA math scores ($F_{(7, 1249)}$ = 652.833, p<.001)[1]. Results for the full model are given in Table 3. When including math indicators in the model, SES was no longer significant, but ethnicity was significant (p<.05). The TN, NS and GPA Math were significant (p<.001). Course total and course type were also significant (p<.05). As indicated by the standardized coefficients, NS was the most influential indicator followed by TN and GPA Math.

Table 3. Regression results for the full model of demographics and all math indicators.

Predictors	**B**	***Beta***	***t***	***p***
Ethnicity	17.975	.034	2.091	.037
SES	10.712	.020	1.319	.187
TN	1.452	.256	11.654	<.001
NS	11.786	.491	18.043	<.001
GPA Math	52.970	.181	10.305	<.001
Course Total	21.238	.027	2.049	.041
Course Type	18.667	.034	1.978	.048

Next, assessment variables and coursework variables were entered in different orders to examine the unique information provided by each set. The left columns of Table 4 show results for a model in which TN and NS were entered in Step 2 and GPA Math, Course Total, and Course Type were entered in Step 3. The two assessments alone accounted for an additional

[1] Multicollinearity, outliers, and assumptions were examined to determine the validity of the full model. Even though there were intercorrelations among some of the predictors, multicollinearity was not a problem. Collinearity statistics showed tolerance values above .1 and variance inflation factors below 10, ranging from 1.046 to 4.311. Cook's measure indicated no influential data points and DfFit values identified only 9 cases as influents, with no pattern in demographics. As for assumptions, relationships among predictors were linear and residuals were normally distributed as indicated by a normal probability plot. A standardized residual plot showed homoscedasticity of residuals.

50.1% of the variance in PSSA scores beyond ethnicity and SES. Coursework variables accounted for another 1.9% of variance.

The right half of Table 4 shows variance accounted for when coursework variables were entered in Step 2 and assessments in Step 3. Although the R^2 Change for GPA Math, Course Total, and Course Type was not as large as Step 2 in the previous model for assessment indicators, the proportion of additional explained variance in PSSA math scores was still quite high (33.4%). In other words, if TN and NS scores were not available, knowing students' math coursework information and their demographics explained 60% of the variance in 11th grade PSSA math scaled scores.

Table 4. Variance in PSSA Scores Accounted for by Math Indicators in Different Orders

Model: Assessments, Coursework			**Model: Coursework, Assessments**		
		R^2 Change			R^2 Change
Step 1	Demographics	.265*	Step 1	Demographics	.265*
Step 2	Assessments	.501*	Step 2	Coursework	.334*
Step 3	Coursework	.019*	Step 3	Assessments	.186*
Total		.785*	Total		.785*

*p<.001

<u>Full Regression Model by Ethnicity</u>

Separate models for each ethnicity were also obtained to determine if the regression on PSSA scores was similar for Black and White students. Because the previous analyses showed that gender was not significantly related to PSSA scores, only SES was entered at Step 1. It explained 4.2% of the PSSA variance in the Black subgroup and 5.9% of the variance in the White subgroup. All math indicators were included in Step 2. Total R^2 was similar for both groups (66.9% for Black and 74.9% for White). Results in Table 5 show similar standardized beta coefficients for the predictors. SES was no longer significant after including the other variables in the model. In both subgroups, the large-scale assessments (TN and NS) were the most influential predictors of PSSA followed by GPA math. For the White subgroup, Course Total and Course Type were not significant. For the Black subgroup, Course Total was significant but the standardized coefficient was quite low.

Table 5. Standardized Regression Coefficients by Ethnicity for Full Regression Equations.

	Standardized Beta Coefficients	
	Black Subgroup	**White Subgroup**
SES	.044	.011
TN	.260***	.267***
NS	.430***	.497***
GPA Math	.246***	.175***
Course Total	.074**	.004
Course Type	.020	.037

**p<.01, ***p<.001

<u>Coursework Regression Models by Ethnicity</u>

Results described above were for a model using all math indicators. The final set of analysis on the subgroups was conducted to examine the influence of coursework variables alone. Total percent of variance explained by SES, GPA Math, Course Total, and Course Type was 53.4% for the White subgroup and slightly lower (41.3%) for the Black subgroup. The standardized beta coefficients in Table 6 indicate that GPA math was the most influential predictor (.538) in the black subgroup. Course Type and SES were also significant but had much smaller coefficients (.183 and .123, respectively). For the White subgroup, GPA math was also the most influential (.460) followed closely by Course Type (.398).

Table 6. Standardized Regression Coefficients by Ethnicity for Coursework Regression Equations.

	Standardized Beta Coefficients	
	Black	**White**
SES	.123***	.070**
GPA Math	.538***	.460***
Course Total	.070	-.032
Course Type	.183***	.398***

**p<.01, ***p<.001

Discussion

This section begins with an interpretation of results for each research question within the context of other literature on test scores, grades, and coursework. Most district personnel are not familiar with the larger research base, and it is beneficial for them to have conversations about how their results fit in with those from other studies. The discussion then turns to viewing the outcomes from a district's perspective and describing how they help identify priorities for schools. Some results lead to further questions of the data which is a natural part of the research process. These additional questions will lead to deeper investigations that can ultimately make a difference in schools and student learning.

Situating Results in the Context of Other Related Research

<u>Relationships Among Math Indicators</u>

The two strongest relationships with PSSA were the TN and NS assessments. It is not surprising that other standardized math tests, even those with different purposes, formats, content, and item types, are strongly related to the state assessment. If students perform well on one test, they tend to perform well on another. In Thacker, Dickinson, and Koger's study (2004), coefficients for TN and PSSA in four districts in Pennsylvania were similarly high.

Two math coursework indicators also had strong relationships with the PSSA. The coefficient for PSSA and GPA math was .672, which is somewhat larger than the coefficients reported by Koger, Thacker, & Dickinson (2004). This might be due to the use of self-reported grades in the state study compared to the use of actual grades in this study. Another possibility is that the state sample was more heterogeneous than the cohort sample and between-school variations may have been large. Results from Willingham et al. (2002) support the latter rationale. They found a coefficient of .63 between total GPA and total NELS scores when using the across-school correlation method. After accounting for between-school variation and several other factors impacting the relationship, the coefficient increased to over .80.

The Course Type indicator was also moderately related to the PSSA. Students who took at least one advanced math course beyond algebra 1, geometry, and algebra 2 tended to score high on the PSSA. The only indicator not significantly related to PSSA was the number of courses taken. Other researchers have also shown that taking more math courses is not necessarily related to higher achievement (Hoffer, 1997; Ma, 2000). Instead, results from national studies indicate that taking advanced mathematics courses is a strong predictor of math achievement, and its effect is still pronounced even after accounting for student background variables (Campbell, Hombo, & Mazzeo, 2000; CEEB, 2001).

Consistency in Relationships Across Demographic Subgroups

Correlational results by student subgroups showed no gender difference in the coefficients. With regard to ethnicity and SES, in most instances there were no significant differences between the Black and White regular lunch groups. The one exception was for Course Type. A stronger relationship between PSSA and Course Type occurred for White regular lunch students compared to Black regular lunch students. When comparing SES, correlations for regular lunch students, regardless of ethnicity, were always higher than correlations for free/reduced lunch students. Most comparisons were statistically significant within the White subgroup but not the Black subgroup.

Other research shows varied results, possibly due to the nature of the samples. In a validity study for the SAT, Young (2004) found that correlations between GPA and SAT were substantially lower for Black and Latino students than for White students. However, in the Willingham et. al. study (2002), correlations between grades and tests were similar. For the ethnic subgroups, interrelationships among the study's variables were consistent. Finally, with respect to international comparisons, a gender gap occurred across countries. In a study of math literacy on the TIMSS (Wilkins, Zembylas, & Travers, 2002), correlations between gender and math literacy were consistently significant. Boys tended to score higher than girls.

Explained Variance in PSSA Math Scores

A high percentage of variance in the PSSA (79%) was explained by the full regression model. Demographics accounted for an initial 27% of the variance, which is consistent with other research on demographics and achievement (e.g., Ma, 2000). NS was the most influential variable, followed by TN and GPA Math. The three coursework variables jointly accounted for 33% of the variance over and above demographics, and the assessments accounted for an additional 19%. Separate regression models for ethnicities were similar in most aspects. Results using the full set of variables explained 67% of variance for Black students and 75% for White students. Standardized coefficients were consistent for the two models, and SES was not significant. Willingham et. al. (2002) also found similarity in regression results for gender and ethnicity subgroups.

However, one difference did occur. When assessments were no longer in the equation, and only coursework indicators and SES were included, the variance explained was 53% for White students, but only 41% for Black students. GPA Math was more influential in the equation for the Black group, whereas Course Type was more influential in the equation for the White group. This result raises a question as to why taking an advanced math course is more influential on PSSA scores for White students than Black students, which leads to the next section.

Interpreting Results from the District's Perspective

This section discusses a few areas of particular interest to the district in their desire to understand and ameliorate inequalities related to student demographics and maximize how well their schools foster student achievement and readiness for success after high school. Some results raise questions for further exploration, other results help to confirm what they suspected based on previous reports of data. Two follow-up studies have occurred since this paper was written (Parke & Keener, 2009; 2011). Selected findings from them are incorporated here.

Advanced Course-Taking

One relationship that stands out among the many examined is between Course Type and PSSA. Correlations between taking an advanced math course and scores on the PSSA were lower for Black students than White students. The correlation was weakest for Black low SES students (.298). Additionally, in the regression analysis, taking an advanced math course was less influential in explaining Black student performance than White student performance. These results cause one to wonder about the upper level math course experiences that are available to Black students, especially low SES. There are several potential hypothesis to explore.

First, not only do advanced courses need to be available to all students, but more importantly the content and instructional strategies must be sound in order for students to have the potential to succeed. Simply enrolling in a high-level course does not necessarily promote math learning and understanding. If students are in an environment that is not positive and does not provide them with worthwhile and meaningful learning experiences, they will not benefit from those courses (Ma & Wilkins, 2007).

Another avenue for exploration is to examine the academic culture, teacher experience, implementation of course curriculum, and grading practices in each high school. Is instructional delivery similar across schools? Do teachers know and understand the math concepts they are teaching? Overall, are some schools better than others at preparing students for success in math? Do all ten high schools offer a range of advanced courses? Are there viable course options for all students? When answering these questions, it would also be helpful to explore reasons for not taking advanced courses, some of which include low math performance in the early grades, low self-confidence, lack of motivation, and lack of encouragement from teachers or parents.

As a follow-up to this study's results, additional analyses were recently conducted within each of the ten high schools (Parke & Keener, 2009). Some findings were not surprising. Two schools that are often touted as top-performing schools in the district had positive results for both Black and White students on all math indicators. These schools had few low SES students, high attendance rates, and few discipline problems. Furthermore, two typically low-performing schools in the district had discouraging results. They had high percentages of low SES students, low attendance rates, and high numbers of disciplinary infractions.

Results were more interesting for other schools. For example, despite several negative contextual factors (majority of students were low SES, mobility rate was high, attendance was low, disciplinary rate was higher than all other schools), something positive seemed to be occurring in one school. Student performance on the state assessment was somewhat above average. The school also had the highest GPA math mean for Black students across the district, one of the highest percentages of Black students taking advanced mathematics across all schools, and one of the lowest percentages of Black students failing math courses. Now the district needs to conduct a qualitative analysis to discover what is occurring in mathematics classrooms in this school.

Gender

Results for gender aligned with other data from the district. They do not need to be as concerned about gender differences as they do about ethnicity or SES differences. Previous reports showed that gender gaps in high school math achievement and coursework were essentially non-existent. However, a recent analysis that disaggregated gender by ethnicity found that equal percentages of White male and White female students took advanced math courses, but a higher percentage of Black females compared to Black males took advanced math (Parke & Keener, 2009). This result will be further investigated within schools.

Information Provided by TN and NS

During the school years analyzed in this study, the district administered two large-scale assessments because they wanted standardized information on student math performance between the 8th grade and 11th administrations of the PSSA. Correlations between TN, NS, and PSSA, technically called validity coefficients, were quite high, especially considering that a whole year passed between taking the tests. Students who scored well on one test scored well on another, which is not uncommon in educational testing. Recently, the district made a decision to no longer use the TN and NS. Instead, they are using a benchmark assessment (4Sight) which gives teachers diagnostic information to analyze and use in making instructional adjustments. Anecdotal reports on how results are used and the impact they have on student learning is positive. Empirical data is now needed to support these claims.

Students Not in the Cohort

The study raised a broader question about the non-cohort students who attended the district high schools for some, but not all, years. District personnel and others hold the belief that most students leave for one of three reasons: 1) families move to a suburban or rural district outside the city limits, 2) families stay in the district but transfer their children to private, religious, charter, or cyber schools, or 3) students drop out of school. Accountability reports show high dropout rates, especially for some schools, but there has not been concrete data on the reasons for exiting.

The district database maintains information on when, where, and why students transfer for the purposes of examining movement of students within and outside the system. In follow-up analyses from this study (Parke & Keener, 2011), a beginning attempt was made to determine when and why students leave. Most non-cohort students left after 9th grade. The average grade in 9th grade math courses was between a "B" and "C" for cohort students versus a "D" for non-cohort students. Tracking students is a complicated process, though. Many non-cohort students had complex withdrawal and reentry patterns that involved moving in and out of the district, attending alternative education centers, and dropping out of school only to return again a few months later. These preliminary results warrant more attention to better understand why enrollment in the district decreases in the high school years.

Final Remarks

Although this study was specific to one school district, there are practical applications that can project out to researchers who investigate mathematics achievement as well as researchers who work with school personnel to help them better utilize student databases. The latter reflects the *process* of conducting a study similar to the one presented here, and the former refers to *knowledge* gained from this study of urban high school students' math performance.

One of the most important steps in the process of helping schools make meaning of their data is to create a clear, specific question that relates to administrators' and teachers' needs. Broad questions such as "what can the data tell us about students' math achievement across our high schools?" will not suffice. Instead, a conversation should take place about the variables to include, the specific sample of students, and the time period upon which to focus. The breadth of data can be overwhelming, but one should resist the urge to include all available variables.

Secondly, try to steer clear of analyses that have already been done and for which everyone knows the answer. For example, some researchers (e,g, Lubienski & Gutierrez, 2008) are now saying that gap analyses that compare mean math scores or percent proficient for one student subgroup versus another are no longer beneficial. This research does not guide further analysis nor does it help in making decisions. In the study described here, the district already knew a gap existed in math scores, so they focused instead on exploring relationships among several math indicators to gain a more in-depth picture of performance in their high schools. Finally, the simplest approach to analysis should be used so that everyone can understand the meaning of the results. As long as it is technically sound and systematically provides an answer to the question, the analysis does not need to be fancy or unnecessarily complex.

A few key outcomes of this study may be of interest to researchers of mathematics educators and school personnel. The number of math courses taken in high school was not related to math scores in any way. However, the type of math course taken was related, and it varied by race. These results helped to set priorities for further analysis in the school district, generating questions such as: Why is taking advanced math scores more influential on math scores for White students than Black students? Are White students learning more in the advanced math courses? Are both subgroups of students equally prepared to take the advanced courses?

Advanced course-taking and grades have been shown to vary across subgroups in a few other research studies in mathematics education (e.g., Riegle-Crumb, 2006). To answer these types of questions in large school districts, analysis by high school could be undertaken. Possibly, teachers are more mathematically experienced and provide higher quality instruction at one school versus another; or the overall school environment and culture at one school might be more positive toward learning and enjoying math than at another school. In smaller districts, additional indicators of math performance can be examined at the classroom level, such as samples of student work and the cognitive level of math discussions. The ultimate goal in these further analyses is to improve upon the teaching and learning process in all math classrooms.

References

A+ Schools. (2007). *The Third Annual Report to the Community on Public School Progress*. A+ Schools: A Community Alliance for Public Education, Pittsburgh, PA.

ACT. (2004). *Issues in College Readiness: The Sensitivity of the ACT to Instruction.* Iowa City, IA: Author.

Brooks-Young, S. (2003). Technology's role in accountability reporting. *School Planning and Management*, *42*(5), p. 42-44.

Byrnes, J. P. (2003). Factors predictive of mathematics achievement in White, Black, and Hispanic 12th graders. *Journal of Educational Psychology, 95*, 316-326.

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance.* Washington, DC: U.S. Department of Education.

College Entrance Examination Board (2001). *2001 profile of college-bound seniors*. New York, NY: Author.

Data Quality Campaign. (2009). *The Next Step: Using Longitudinal Data Systems to Improve Student Success*. Author.

Enomoto, E. K. & Conley, S. (2007). Harnessing technology for school accountability: A case study of implementation. *Planning and Change, 38*(3/4), p. 164-180.

Goldschmidt, P., Martinez, J. F., Niemi, D., & Baker, E. L. (2007). Relationships among measures as empirical evidence of validity: Incorporating multiple indicators of achievement and school context. *Educational Assessment, 12*(3&4), 239-266.

Hall, C. W., Davis, N. B., Bolen, L. M., & Chia, R. (1999). Gender and racial differences in mathematical performance. *Journal of Social Psychology, 139*(6), 677-89.

Hoffer, T. B. (1997). High school graduation requirements: Effects on dropping out and student achievement. *Teachers College Record*, *98*, 584-607.

Koger, M. E., Thacker, A.A., & Dickinson, E.R. (2004). *Relationships among the Pennsylvania system of school assessment (PSSA) scores, SAT scores and self-reported high school grades for the classes of 2002 and 2003*. HumRRO Report No. FR-04-26.

Lubienski, S. T. & Gutierrez, R. (2008). Bridging the gaps in perspectives on equity in mathematics education. *Journal for Research in Mathematics Education*, *39*(4), p. 365-371.

Ma, X. (2000). A longitudinal assessment of antecedent coursework in mathematics on subsequent mathematical attainment. *Journal of Educational Research, 94*, 16-28.

Ma, X. & Wilkins, J. L. M. (2007). Mathematics coursework regulates growth in mathematics achievement. *Journal of Research in Mathematics Education, 38*(3), 230-257.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM.

Parke, C. S. (2006). *Student Attendance and Mobility and the Effects on Student Achievement in Mathematics and Reading*. Technical Report #06-12-01. Pittsburgh, PA: A+ Schools: A Community Alliance for Public Education.

Parke, C. S. (March 2008). *High School Mathematics Achievement for a Cohort of Students (2002 to 2005): Large-Scale Assessments and Math Coursework.* Technical Report #08-03-01. Pittsburgh, PA: A+ Schools: A Community Alliance for Public Education.

Parke, C. S. (2009). *Mathematics Achievement for Cohort and Non-Cohort High School Students (2002 to 2006): Large-Scale Assessments and Math Coursework*. Technical Report #09-05-01. Pittsburgh, PA: A+ Schools: A Community Alliance for Public Education.

Parke, C. S., & Kanyongo, G. (In press). Student attendance, mobility, and mathematics achievement in an urban school district. *Journal of Educational Research*.

Parke, C. S. & Keener, D. (2009). *Analyzing math assessment scores, math coursework, and contextual factors for high schools in an urban district*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Parke, C. S., & Keener, D. (2011). Cohort versus non-cohort high school students' math performance: Achievement test scores and coursework. Educational Research Quarterly, 35(2), p. 3-22.

Pennsylvania Department of Education. (2005). *Technical Report for the Pennsylvania System of School Assessment: 2005 Reading and Mathematics.* Data Recognition Corporation.

Riegle-Crumb, C. (2006). The path through math: Course-taking trajectories and student performance at the intersection of gender and race/ethnicity. *American Journal of Education*, *113*(1), 101-122.

Schmoker, M. (2008). Measuring what matters. *Educational Leadership*, p. 70-74.

Smith, J. B. (1996). Does an extra year make any difference? The impact of early access to algebra on long-term gains in mathematics achievement. *Educational Evaluation and Policy Analysis*, *18*, 141-153.

Thacker, A. A. & Dickinson, E. R. (2004). *Item Content and Difficulty Mapping by Form and Item Type for the 2001-2003 Pennsylvania System of School Assessment (PSSA).* Alexandria, VA: Human Resource Research Organization.

Thacker, A. A., Dickinson, E. R., & Koger, M. E. (2004). *Relationships Among the Pennsylvania System of School Assessment (PSSA) and Other Commonly Administered Assessments*. Alexandria, VA: Human Resource Research Organization.

Wilkins, J. L. M, & Ma, X. (2002). Predicting student growth in mathematical content knowledge. *Journal of Educational Research, 95(5)*, 288-298.

Wilkins, J. L., Zembylas, M., & Travers, K. J. (2002). Investigating correlates of mathematics and science literacy in the final year of secondary school. In D. F. Robataille & A. E. Beaton (Eds.) *Secondary Analysis of the TIMSS Results: A Synthesis of Current Research* (p. 291-316. Boston, MA: Kluwer Academic Publishers.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*, 1-37.

Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In R. Zwick (Ed.), *Rethinking the SAT: The Future of Standardized Testing in University Admissions* (pp. 289-301). New York: Routledge/Falmer.

Zwick, R. & Green, J. G. (2007). New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. Journal of Educational Measurement, 44(1), p. 23-45.

Zwick, R. & Schlemer, L. (2004). SAT validity for linguistic minorities at the University of California, Santa Barbara. *Educational Measurement: Issues and Practice, 23*(1), 6-16.

Zwick, R. & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. American Educational Research Journal, 42(3), 439-464.