



# Summary: Evaluation of Mathematics with Meaning and Textual Power through Student Achievement Analysis

Christine Leow • Kerstin Carlson Le Floch



AMERICAN  
INSTITUTES  
FOR RESEARCH®

1000 Thomas Jefferson Street, NW | Washington, D.C. 20007-3835 | 202.403.5000 | [www.air.org](http://www.air.org)

---

## Table of Contents

Perspective .....	1
Purposes and Objectives .....	2
Research Questions.....	2
Research Design and Methodology .....	2
Data Sources .....	2
Study Design.....	3
Selection of Comparison Classes.....	3
Background Variables Considered .....	4
Analytical Method .....	4
Construction of Separate Analyses .....	5
2002-2003 Results .....	6
2003-2004 Results .....	7
Discussion.....	8
What Makes a Mathematics with Meaning or Textual Power Class Different?.....	9
Conclusion .....	9

---

## PERSPECTIVE

Student achievement depends on an array of factors within and beyond the classroom environment. Even within classrooms, a complex interplay of factors affects achievement. These factors include not only the curriculum but also teaching practices. Reform in the classroom entails both new teaching materials and new instructional strategies that are associated with the curriculum. In this era of accountability, we need to know if reforms actually work and are not based on ideals that are challenging to implement, and demonstrate no measurable gains. Districts adopting reform programs need to know if they have the promise to promote improved student achievement.

In this paper, we are interested in two associated pilot programs, consisting of professional development, instructional strategies, and curricular materials and how these programs impact student achievement. Specifically, we examine the effectiveness of the Mathematics with Meaning and Textual Power pilot programs developed by the College Board. Mathematics with Meaning is used in mathematics classes while Textual Power is used in English classes. These programs are separated into middle school and high school. The high school pilot programs were piloted in 2000-2001 while the middle school programs were piloted in 2001-2002. However, these programs do not constitute full curriculum but are supplementary programs that include professional development, instructional strategies, and planned classroom materials.

The main premise of both programs is that students learn within meaningful contexts. Instead of relying on traditional teaching straight from the textbooks, these programs emphasize the use of materials and real-life situations to teach lessons. Theoretically such programs, that are based on the premise that learning with meaning, should work because it follows one of the guiding principals of constructivism where children learn best within meaningful context. If teachers who attended the professional development fully implemented the programs by successfully engaging their students in the academic matter, an increase in comprehension, interest, and achievement of students is possible. However, how do these reform-based supplementary programs hold up in practice? To investigate the link between Mathematics with Meaning and Textual Power and student achievement, we need to conduct a well-designed study that matches students exposed to the programs with comparable students not exposed to the programs. It is only through comparison with a well-matched group that we are able to draw valid conclusion about the programs' effectiveness. Achievement on state assessments as an outcome measure is the focus of this study because it is one of the main, if not singular, emphases in school and district accountability.

---

## **PURPOSES AND OBJECTIVES**

There are two purposes in this paper: To evaluate Mathematics with Meaning and Textual Power and to present a valid method for the evaluation that is based on well-matched comparisons. The objectives are framed explicitly with the following research questions:

### **Research Questions**

- Do students in the classrooms using Mathematics with Meaning and Textual Power show improvements in achievement compared with students in similar classrooms not using the programs?
- How should the comparison classrooms be selected so that they are comparable to the treatment and hence, make a valid conclusion about the effectiveness of the programs?

## **RESEARCH DESIGN AND METHODOLOGY**

### **Data Sources**

In this study, we had obtained student achievement data from two school districts for the academic years 2002-2003 and 2003-2004. These two districts had sent some of their teachers to the College Board's Mathematics with Meaning and Textual Power professional development. However, not all teachers who received training from the College Board actually implemented Mathematics with Meaning or Textual Power in their classes. To be sure that our treatment classes were truly those in which teachers were using the pilot programs, we relied on several sources of data, including our own site visits and correspondence with district administrators and teachers. In one district, an administrator personally contacted every trained teacher within his district to obtain course IDs for each class in which they were using the pilot program materials in that particular year. With this information, it is thus possible to identify those classes whose teachers went for the professional development and implemented Mathematics with Meaning or Textual Power in the classes. These classes were designated as treatment classes. Similarly, the researcher contacted every mathematics and English teacher that went for the professional development in the other district through phone or email to determine which classes could be designated as treatment classes.

---

## Study Design

To examine the association of the use of College Board materials with student achievement, we classified classes into two groups—the treatment group and the comparison group. Students in treatment classes were those whose teachers underwent professional development and used Mathematics with Meaning or Textual Power. They were compared to a matched group of comparison students from the same district whose teachers did not attend College Board professional development and therefore did not use the programs. This made it possible to calculate the differential achievement gap between College Board classes and comparison classes through a pretest-posttest with matched comparison group research design in the analyses. The comparison classes were matched to the treatment classes on three crucial variables—prior achievement, grade level, and course type.

To prepare for the analyses, we linked student achievement data from his/her score the previous year so that we could determine each student’s prior achievement, which we needed to use as a control in the analysis as well as during matching. To link student achievement data, we assigned unique identification numbers to all students, classrooms, teachers, and schools. For the analyses, students and teachers were identified by their unique identification number and not by their names to ensure confidentiality. With these identification numbers, and another variable that indicated the use of Mathematics with Meaning or Textual Power, we grouped students by treatment versus comparison classes and then compared the achievement gap.

## Selection of Comparison Classes

In the districts that were used in this study, the majority of the teachers who received the College Board training and were implementing the programs usually use them in all the classes they teach. However, a few teachers were not implementing the programs in all their classes. Hence, we see an implementation design at two levels—some at the teacher level (for teachers who used the materials in all their classes) and some at the class level (for teachers who did not use the materials in all their classes). Because of this mixed design in implementation, we chose the lower unit of implementation and matched at that level, which is the class level. In addition, the same teacher might be teaching different types of classes and it would be inappropriate to combine different types of classes taught by the same teacher and analyze at the teacher level. For example, a teacher could be teaching basic algebra and higher level algebra and it would be inappropriate to combine the two classes to do a teacher level analysis.

---

Comparison classes were first matched exactly to the treatment classes on grade level and course type. Next, we matched them on prior achievement by selecting, a priori, a matching distance (caliper) on the achievement score and selected those comparison classes whose prior achievement mean fell within the caliper of each of the treatment class' prior achievement mean. We determined the matching caliper by examining the standard deviations among the classes within each grade level and course type. If the standard deviation was high, we used a larger caliper to reflect the greater variation. A general rule was to have the caliper a quarter to a half of the standard deviation among the classes.

## Background Variables Considered

There are also other student and teacher background variables that could affect student achievement, including sex, race, free or reduced-price lunch participation, individualized educational plan (IEP) status, Limited English Proficiency (LEP) status, and teacher certification. We considered some of these variables in our analyses, but we were limited by the depth of the data files that were maintained by the participating districts. Although the number of background variables that were available in the different districts differed, students' prior achievement scores (the most crucial type of control), grade level, and type of course were fully available in all the districts under study.

## Analytical Method

Hierarchical Linear Modeling (HLM) was used for analyses of the association between Mathematics with Meaning and Textual Power and student achievement. Below we provide the general model used for one district and one subject. Owing to the nature of the data, there were two levels of data—student and class.

To account for student contextual factors that might be associated with student achievement, such as sex, race, free or reduced-price lunch participation, individualized educational plan (IEP) status, Limited English Proficiency (LEP) status, and grade level, we specify the student-level model as:

$$Ach_{ij} = \beta_{0j} + \beta_1 Ach(n-1)_{ij} + \beta_2 demos_{ij} + r_{ij} \quad (1)$$

$Ach_{ij}$  is a function of the previous test score and demographic characteristics (sex, race, free or reduced-price lunch participation, IEP status, LEP status, and grade level), denoted by  $demos_{ij}$ .

---

Because the College Board pilot programs were implemented at the class level, the class-level model is specified as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \textit{treatmentgroup}_j + \gamma_{02} \textit{course}_j + u_{0j} \quad (2)$$

Equation (2) depicts the class-level model.  $\beta_{0j}$  represents the intercept from the student-level model in Equation (1). As shown in Equation (2), the difference of achievement mean scores between classes using and not using Mathematics with Meaning or Textual Power is determined by the variable  $\textit{treatmentgroup}_j$ , which was coded as 1 if a class used the College Board materials and 0 if it did not. When the coefficient of  $\textit{treatmentgroup}_j$  is positive and significant, this indicates that the use of Mathematics with Meaning or Textual Power appeared to be linked to higher student achievement. The term  $\textit{course}_j$  was used as a general term to denote the various types of courses in which the students in the analyses were enrolled. For example, a student that was enrolled in a basic algebra class would be dummy coded as 1 for that course.

## Construction of Separate Analyses

Analysis was performed separately for the two school districts and separately by subject (mathematics and English). As we had sufficient students from middle and high school, we were also able to do the analyses separately by middle and high school. Hence, we have seven separate analyses, minus District B Mathematics in Middle School where Mathematics with Meaning was not implemented.

1. District A, Mathematics, Middle School
2. District A, English, Middle School
3. District A, Mathematics, High School
4. District A, English, High School
5. District B, English, Middle School
6. District B, Mathematics, High School
7. District B, English, High School

In addition, it should be noted that analyses for the academic year 2003-2004 were done separately from academic year 2002-2003. This is because student IDs changed from one year to the next and hence we were not able to track each student's progress from 2002 to 2003 to 2004.

---

## 2002-2003 Results

Significant findings were found at the high school level in District A, in both English and mathematics. The results were significant in mathematics at the 0.05 level and significant in English at the 0.01 level. It was not found to be significant in District A middle schools and in the whole of District B. For mathematics, students in classrooms using the College Board materials scored, on average, 15 points higher on their 2003 achievement test. In English, the treatment students scored, on average, 46 points higher. The achievement data for both subjects are expressed in scaled scores that span both middle and high school grades, with an expected mean of 1,800. Other than prior achievement (where it was positively significant to student achievement across all seven analyses), there appeared to be no consistent pattern of significance on other student and teacher variables used in the analyses.

What does the difference between the treatment and comparison mean? The chart below shows the entire State mean gains from one grade to the next in 2002 for mathematics and English. For example, in mathematics, students on average gained 98 points from grade 9 to 10, and 156 points in English. The estimated treatment effect of the Mathematics with Meaning in high school mathematics was 15, and the effect for English was 46. The size of the treatment effect is substantial relative to the typical gain. A student who might be expected to gain 156 points in English without the treatment, for example, might be expected to gain 202 ( $156 + 46$ ) with the treatment.

### State Mean Gain for the Developmental Scaled Score in 2002-2003

	Mean Developmental Scale Score Gain	
	<i>Mathematics</i>	<i>Reading</i>
Gain from grade 6 to 7	110	89
Gain from grade 7 to 8	103	128
Gain from grade 8 to 9	34	-28
Gain from grade 9 to 10	98	156

These results provide preliminary evidence that Mathematics with Meaning and Textual Power can be associated with higher student achievement, as measured in a state assessment in a high school district. One likely reason for the significance in high schools is because the pilot programs were first implemented in high schools a year earlier than in middle schools. This evaluation took place at the end of the second school year of implementation in high schools and at the end of the first year of implementation in middle schools.



---

For the significant findings found in District A and not District B, it could likely be due to broader participation of all the education stakeholders at District A who agreed to adopt Mathematics with Meaning and Textual Power. Furthermore, there appeared to be a wider network of support for teachers and greater access to program materials in District A. These were evidenced in a survey sent to teachers during the first year of evaluation.

## **2003-2004 Results**

In the second year of evaluation, it was found that Mathematics with Meaning and Textual Power were associated with higher student achievement in District A at the middle school in both subjects, and at the high school level in mathematics only. There were many missing data points in District A high school English. Hence, it is difficult to draw much conclusion about the insignificant results found in District A high school English. Mathematics with Meaning and Textual Power were not found to be associated with higher student achievement in District B.

In District A middle school, students in classes using Mathematics with Meaning scored on average 22.7 points higher than students in classes not using the College Board materials, which was found to be statistically significant at the 0.05 level. For District A middle school English, results showed that students in classes using Textual Power were achieving on average 20.9 points higher than student in classes not using Textual Power, which was statistically significant at the 0.01 level.

To further understand the significance of the treatment effect, the chart below displays the state's mean gain for the developmental scale score from one grade to another in 2003-2004. Thus, for a significant finding of a difference of 22.7 in mathematics, it was almost an entire grade gain when one moved from grade 5 to grade 6, which means that the treatment student can be achieving at a grade level higher. The gain was about a fifth of a grade gain when one moved from either grade 6 to 7 or from grade 7 to 8, with the treatment student achieving at a fifth of a grade higher at these grade levels. Similarly, for English, an average difference of 20.9 points between treatment student and comparison student was about a fifth of a grade gain, from grade 5 to grade 6 or from grade 6 to grade 7 or from grade 7 to grade 8.

In District A high school, the variable *treatment group* revealed the results to be significant at the 0.05 level. Thus, students in classes using Mathematics with Meaning scored 17.0 points higher on their achievement than students who were in classes not using Mathematics with Meaning.

Again, to further understand the significance of using Mathematics with Meaning, the table below shows the state's mean gain for the Developmental Scale Score as students moved from grade

---

8 to grade 9 from 2003 to 2004. Thus, with a significant difference of 17 between treatment and comparison groups, a Mathematics with Meaning student was achieving between a quarter to half a grade higher than a comparison student.

### **State's Mean Gain for the Developmental Scale Score from 2003 to 2004**

	Mean Developmental Scale Score Gain	
	<i>Mathematics</i>	<i>Reading</i>
Gain from grade 5 to 6	30	94
Gain from grade 6 to 7	118	91
Gain from grade 7 to 8	111	111
Gain from grade 8 to 9	47	-12

There were significant findings in District A for two years though there were none at all in District B. Many factors could account for the insignificant findings in District B. Most notably, the sample sizes for District B were unusually small because of missing data. For this reason, we are less confident of these analyses. Because the district gave no explicit reason for the missing data, performing imputation was not an appropriate option. Based on our interview and other site visit data, Mathematics with Meaning and Textual Power appeared to be adopted at the school level in District B. Thus, teachers at the schools have limited support, especially from the district, if they decided to use Mathematics with Meaning and Textual Power.

## **DISCUSSION**

This study is useful on many dimensions. First, it provides tangible, scientifically-based evidence regarding the association of a reform-based, supplementary program with higher student achievement. Second, a sound methodology was used for this evaluation through the process of selecting well-matched comparisons within the same district.

Our student achievement findings were buttressed by classroom observations. Classroom observations noted positive key differences found in Mathematics with Meaning and Textual Power classrooms. These include number of instructional activities, nature of these activities, level of student engagement, and student behavior. Hence, what exactly distinguishes a College Board class from a comparison class?

---

## **What Makes a Mathematics with Meaning or Textual Power Class Different?**

Teachers that went to the Mathematics with Meaning professional development used more manipulatives. Classroom observations also showed that there was an emphasis on Problem Solving, Reasoning and Proof, Communication, Connections, and Representation strategies in Mathematics with Meaning classes compared to comparison classes. These strategies are skills that are reflective of the NCTM standards. Mathematics with Meaning teachers tended to use more instructional strategies that were more interactive than the comparison teachers and they spent more time leading and supporting student work.

Teachers in Textual Power classes used more text-based materials such as fiction, drama, and poetry than comparison classes. They tended to use strategies such as Making Meaning from Text that reflected students' interpretation of the text while comparison teachers tended to use more Practice and Drill strategies that reflected more rote learning. For both Mathematics with Meaning and Textual Power classes, it was also observed that there tended to be a higher rate of student on-task participation than comparison classes.

## **CONCLUSION**

Thus, this evaluation appeared to point to the effectiveness of Mathematics with Meaning and Textual Power. However, there is a major caveat to this evaluation. Although there were some background variables that were available to control for, these were limited. Hence, caution should be taken in extending the results to different settings.