Empirical Education

RESEARCH REPORT

Comparative Effectiveness of the Texas Instruments TI-Navigator™:

Year 2 Report of Randomized Experiments in the East Side Union High School District and San Diego Unified School District

Brandon Hoshiko
Andrew Jaciw
Boya Ma
Gloria I. Miller
Xin Wei
Empirical Education Inc.

December 2007

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 210
Palo Alto, CA 94306
(650) 328-1734

Acknowledgements

We are grateful to the people in the East Side Union High School District and the San Diego Unified School District for their assistance and cooperation in conducting this research. The research was sponsored by Texas Instruments, which provided Empirical Education Inc. with independence in reporting the results.

About Empirical Education Inc.

Empirical Education Inc. was founded to help school districts, publishers, and the R&D community assess new or proposed instructional and professional development programs through scientifically based pilot implementations. The company draws on the expertise of world-class researchers and methodologists assuring that the research is objective and takes advantage of current best practice in rigorous experimental design and statistical analysis. The company's findings let educators quantify the value of programs and help them partner with providers to implement those most effective for their students.

Executive Summary

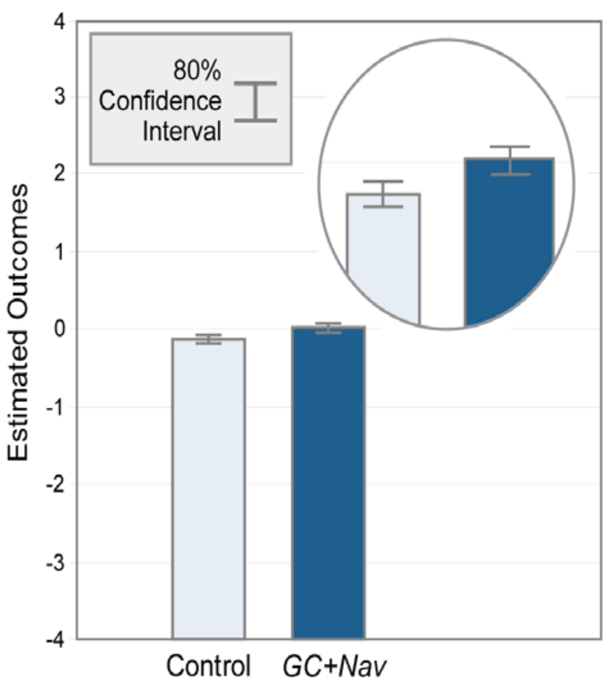
Introduction

We sought evidence of the effectiveness of the TI-Navigator classroom networking system for the second year of a two-year research study of Texas Instruments classroom technology. This randomized control trial compared Algebra I and Geometry instruction using the TI-Navigator system, which includes the TI-84 Silver Edition graphing calculator, to instruction with graphing calculators alone. The technologies did not provide a separate math curriculum, but did include curricular materials that specified calculator-based activities. The outcomes of interest are the student test scores in Algebra I and Geometry. For both subjects, we had two sets of scores: the Northwest Evaluation Association (NWEA) End of Course tests and the California Standards Test (CST).

We asked whether students in classrooms with access to the TI-Navigator system and training achieve higher scores than students in classrooms receiving only graphing calculators and training, whether TI-Navigator has a differential impact for students with various incoming math achievement levels, and whether the impact depends on gender or English proficiency. We researched these questions for classes in two urban school districts in San Jose and San Diego, California. East Side is a high school district within San Jose of about 25,000. San Diego City Schools is a K-12 district of about 135,000. In both cases about 28% are English learners.

Findings

For the most part, the experiment could not discern an impact as a result of providing the equipment and training for TI-Navigator. As shown in the figure below, we found a modest effect for Geometry achievement using the NWEA End of Course Geometry test. This figure shows the outcome measure in standardized units. However, this impact was not reflected in CST Geometry scores.



Impact on NWEA Geometry Achievement

In Algebra, while we found no overall difference, there was some evidence of a small negative impact for students scoring “below basic” on the CST and, holding pretest score constant, for English proficient students. The results of the NWEA End of Course Algebra I test did not reflect those same results.

Implementation must be considered in interpreting these findings. Our surveys and observations make clear that this implementation was not a fair test of the difference TI-Navigator might make if used more extensively. Of the 19 teachers originally assigned to the treatment group, about half did not use the system at all for instruction. Of the remaining nine teachers, only three could be considered “Comprehensive-Implementers.” Of those three, only one used TI-Navigator daily. Technical glitches deterred many from using the system after previous failed attempts. Overall use may have been constrained by the fact that California prohibits calculator use on state tests.

Our results also must be qualified by the fact that, while finding differences on one test, we did not find differences on the other test. The significant amount of attrition, both at the teacher and student levels, although not believed to be associated with the program being tested, raises issues about generalizability. For example, it is clear that in both experimental conditions, lower scoring students were significantly more likely to not have posttests, indicating that our findings are not applicable to the lowest scoring students in these districts.

Overall, we found that the TI-Navigator affected the average number of minutes the technology was used. The teachers with TI-Navigator reported using the technology about 15 minutes more per week per class period than teachers without. Future exploratory analyses may prove useful in suggesting whether extent of usage can account for student outcomes. In particular, since TI-Navigator resulted in greater technology use, examining the correlation between technology use and achievement may suggest a mechanism by which TI-Navigator could be effective. Future studies of TI-Navigator will benefit from greater support for implementation. We also recommend continuing to include Geometry in the topics to which TI-Navigator is applied, since the positive result found in this experiment should be replicated.

Design and Analysis

In the first year of this two-year experiment, we used a matched pair design to randomly assign 44 teachers to use graphing calculators with their existing math curriculum or to conduct “business as usual” in the classroom. In this second year, teachers kept their random assignments, the original graphing calculator group receiving TI-Navigator and the original control group receiving graphing calculators. The technologies were intended to be integrated with the school’s standard Algebra I and Geometry curricula.

All teachers participating in the study received TI graphing calculators, which have several features that can be used in Algebra I and Geometry classrooms. Program group teachers also received the *TI-Navigator*™ 3.0 system, which is designed to work with the TI graphing calculators and adds two capabilities: 1) wireless communication between students’ graphing calculators and the teacher’s PC computer and 2) activity center, quick polling, and screen capture activities. Separate, three day, professional development was provided for both sets of technology, the TI-Navigator system with graphing calculators and the graphing calculator alone. TI also provided all study teachers a standard notebook computer with a calculator emulator, a data projector, and calculator-based ranger units.

For this experiment, we randomized teachers in approximately equal numbers to the TI-Navigator and control groups. Because results from the experiment’s first year suggested a differential impact by English language learner status, we examined this moderator to determine whether the effect can be replicated. We also examined gender as a potential moderator.

The data for this study consist of student outcomes, demographics, and classroom observations. In addition to conducting formal and informal teacher interviews, we also collected 15 web-based survey responses from all participating teachers in each group. We retrieved TI-Navigator system log files from pilot classroom computers to confirm TI-Navigator use.

We designed the experiment described in this report to provide useful information to the participating school districts. Because we were testing a specific implementation of TI-Navigator in a particular setting, we caution readers about generalizing the results to districts with different populations, resources, and other relevant conditions. Although our results cannot be used as definitive evidence of the value of TI-Navigator*,* the areas of positive findings lead us to recommend that schools planning to implement TI-Navigator provide adequate support, both technically and educationally, while rolling out implementation in a manner that allows for continued tracking of student achievement gains.

Comparative Effectiveness of the Texas Instruments TI-Navigator™:

Year 2 Report of Randomized Experiments in the East Side Union High School District and San Diego Unified School District

Table of Contents

Introduction

This is the final report on the second year (Phase II) of a two-year research study of Texas Instruments classroom technology. We report here on research aimed at producing evidence of the effectiveness of the TI-Navigator classroom networking system. This randomized experiment or randomized control trial (RCT) compared Algebra I and Geometry instruction using the TI-Navigator system, which includes the TI-84 Silver Edition graphing calculator to instruction with graphing calculators alone. The technologies were intended to be implemented after summer trainings for the TI-Navigator system and graphing calculators. The technologies were also intended to be used with the school districts' existing math curriculums and did not provide a separate math curriculum, but did include curricular materials that specified calculator-based activities.

The specific research questions addressed are:

- Do students in classrooms with access to the TI-Navigator system (including graphing calculators) and professional development achieve higher scores in Algebra and Geometry than students in classrooms that only received graphing calculators and graphing calculator training?
- Does TI-Navigator have a differential impact for students with different incoming math achievement levels? Does the impact depend on gender? Does it depend on English proficiency?

We were able to ask these questions for both Algebra and Geometry classes in two urban California school districts.

Description of Phase I

From a pool of volunteers, we randomly assigned teachers to those using the TI-84 Silver Edition graphing calculator (*GC* group*)* and those in the "business as usual" condition (control group). Randomization was stratified according to whether the teacher taught Algebra I or Geometry, years teaching experience, and the school's Title I status and English learner population. Altogether, the experiment involved 33 teachers (17 *GC* and 16 control group teachers), 80 classes, and 1,201 students.

In the fall of 2005, Texas Instruments provided training, equipment, and activities for teachers participating in this experiment. Phase I is divided into two parts. Two sets of materials were deployed: the TI-84 Silver Edition Plus graphing calculator system technology (during the first semester) and the TI-Navigator™ 2.0 system (second semester, to a subset of intervention teachers). After one semester, we divided the *GC* group into teachers who continued to use the calculators and those using the calculators with TI-Navigator 2.0 (*GC + Nav*) as a further enhancement of the instructional capabilities of the technology. In this way we apportioned the first year into two semester-long experiments which also combine to form a year-long experiment. Besides allowing us to get an initial understanding of the enhancement, which was to be used systematically in the second phase of the experiment, this also allowed us to put all the elements into practice in the first year so that implementations could be refined as necessary for Phase II. These classrooms became what we now call the *GC+Nav (A)* group. Thus, at the conclusion of the first phase, there were three groups of classrooms in the study, the *GC+Nav (A)* group, the *GC* group, and the control group.

We examined the comparative effectiveness of classrooms using *GC+Nav (A)* group and the *GC* group in contrast to the control group. For the results of this first phase, please see Phase I Final Report. (Miller, Jaciw, Hoshiko, Vu & Wei, 2007).

Description of Phase II

The same teachers continued into the second phase of the study in which the former *GC* group became the TI-Navigator group (*GC+Nav*) while the former control group was provided graphing calculators and training, becoming the new control group, but now using the graphing calculators. In the second year of the study, Phase II, data collection continued through the 2006-2007 academic year. In August 2006, **all** classrooms in the previous *GC* group have the TI-Navigator system. This group is composed of two subsets; one group received the TI-Navigator system and training mid-way through the first year (January 2006) and the other group received the TI-Navigator system and training at the start of the second year (August 2006). We refer to these two groups as *GC+Nav (A)* and *GC+Nav (B)* respectively*.* The small number of classrooms that used the TI-Navigator technology in each of the two groups prevents us from drawing definitive comparisons between the *GC+Nav (A)* and *GC+Nav (B)* groups, but the distinction did give us an initial indication of longer term usage. In reporting the second year experiment, the *GC+Nav (A)* group and *GC+Nav (B)* group will be combined and referred to as *GC+Nav*. Where a distinction between the two groups is needed, we will refer to them individually as *GC+Nav (A)* and *GC+Nav (B).*

For the experiment being reported here, the previous control group teachers received training and graphing calculators for their eligible classes. We will continue to refer to this group as the control group.

Random assignment to experimental conditions does not assure that we can generalize the results beyond the districts where the research was conducted. We designed our study to provide useful information to support local decisions that take into account the specifics of district characteristics and their implementation of the program. The results should not be considered to apply to school districts with practices and populations different from those in this experiment. The report provides a rich description of the conditions of implementation in order to assist the district in strengthening its program and to provide the reader with an understanding of the context for our findings.

Methods

Our experiment is a comparison of outcomes for classrooms using TI-Navigator with graphing calculators (*GC+Nav* group) and classrooms using graphing calculators alone (control group). The outcomes of interest are the student test scores in Algebra I and Geometry. For both subjects, we had two sets of scores: the Northwest Evaluation Association (NWEA) End of Course tests and the California Standards Test (CST)

This section details the methods used to assess the size of the difference in test scores between the *GC+Nav* program and control groups (within confidence limits set by the available sample size) and whether the introduction of TI-Navigator was responsible for those differences. Additionally, we discuss the methods used to examine the implementation of the intervention. We begin with a description and rationale for the experimental design and go on to describe the intervention, the research sites, the sources of data, the composition of the experimental groups, and finally the statistical methods used to generate our conclusions about the impact of *GC+Nav*.

Experimental Design

With experiments we usually randomize an available sample of cases. Generalization is left to heuristic arguments, which include a comparison of the characteristics of the sample with that of the population of interest (e.g., the whole district). Though we don't have the luxury of selecting a random sample of cases from the population before assigning cases to treatment or control, our results need to express the fact that our sample is just a select group of cases, and the results we get would change if a new sample of teachers or students was selected into the experiment by whatever mechanism. The design of the experiment is based on our best understanding of the amount of variability that we expect due to re-sampling, where our intention is to limit the effect of this "noise" in order to detect the stable signal (the effect), if it exists. There is always a level of uncertainty and an associated level of imprecision. We think of the uncertainty as related to the likelihood that we would get a different result if we took a new sample of students or of teachers from the same larger population. Our design attempts to efficiently deploy the available resources to reduce uncertainty and improve precision, in other words, to reduce the likelihood that we would get a different result if we tried the experiment again.

An upfront effort to fully specify a design or plan for the experiment pays off in two ways. First, we identify, before seeing the outcomes, where we expect to see an impact and what factors we expect will moderate the impact. In other words, we specify the research questions in advance. In this way, we avoid fishing for results in the data, a process that can lead to mistaking chance differences for differences that are probably important as a basis for decisions. Because some effects will be big simply by chance, “mining” the data in this way can capitalize on chance. We can still explore the data after the fact, but this is useful mainly for generating ideas about how the new program worked, that is, as hypothesis-generating efforts for motivating future study, rather than as efforts from which we make firm conclusions from our existing study.

Second, an experimental design will include a determination of how large the study should be in terms of students, teachers, and schools in order to get to the desired level of confidence in the results. In the planning stage of the experiment we calculate either how many cases we need to detect a specific sized difference between the program and control groups, or how big a difference we can detect given the sample size that is available. Technically this is called a power analysis. We will explain how many aspects of design determine the size of the experiment.

Design Features to Address the Research Questions

How the Sample was Identified

How the participants for the study are chosen largely determines how widely the results can be generalized. In this case, Texas Instrument personnel initially introduced members to our research team to the East Side Union High School and San Diego Unified School districts as sites interested in the TI-Navigator with graphing calculator technology and willing to conduct an RCT experiment with a sample of their classes.

We initially met with district staff members from interested sites to explain the details and procedures of the study. Our district point of contacts identified teachers who did not have TI-Navigator experience and were expected to teach Algebra I or Geometry classes in the fall. These teachers were then invited to after-school meetings. Only schools in which eligible teachers agreed to participate were included in the study. The initial meeting for the research experiment in ESUHSD occurred on May 31, 2005 with 22 teachers who teach Algebra I and/or Geometry. The initial meeting for the research experiment in the SDUSD took place on June 15, 2005 with 22 teachers who teach Algebra I and/or Geometry.

Randomization

Since we want to know the impact of the Texas Instruments’ technologies, we have to isolate its impact from all the other factors that might make a difference for how or what teachers and students do. We want to answer whether TI-Navigator caused a difference. Randomization ensures that, on average, characteristics other than the program that affect the outcome are equally distributed between program and control. This distribution prevents us from confusing the program’s effects with some other factors, technically called “confounders,” that, because they also affect the outcome, would lead to bias if they are unevenly distributed between the groups. For example, randomization helps to ensure that more experienced teachers are not selectively assigned to the program or control group.

Organizational Levels Considered in the Experiment

This research works within the organization of schools by not disrupting the existing hierarchy in which students are grouped under teachers who belong to schools. The level in the hierarchy at which we conduct the randomization is generally determined on the basis of the kind of program being tested. School-wide reforms call for a school-level randomization, while a professional development program can use a teacher-level randomization. Generally, we attempt to identify the lowest level at which the intervention can be implemented without unduly disrupting normal collaboration and without inviting sharing or “contamination” between control and program units. For this experiment, we randomized teachers in approximately equal numbers to the *GC+Nav* and control groups. The outcome measures are student-level test scores in Algebra and in

Geometry. Because teachers instead of students were assigned to *GC+Nav* or control, this kind of experiment is often called a "group randomized trial."

What Factors May Moderate the Impact of TI-Navigator?

Our design allows us to consider the extent to which TI-Navigator is more effective for students of different English language status and for male or female students. These are variables that are measured before the experiment starts, and that we have reason to believe will affect the strength of the effect of TI-Navigator. Technically, these are called potential moderators because they may moderate the impact of TI-Navigator. During analysis we measure the strength of the interaction between each moderator and the TI-Navigator effect; that is, we measure whether the effect of TI-Navigator changes as the level of the moderator changes.

Because the results from Phase I of the experiment suggested a differential impact for students of different English language learner status, we examine this moderator to determine whether the same effect can be replicated. We also examine gender as a potential student-level moderator.

How Large a Sample Do We Need?

A process called power analysis was used to plan the number of teachers that the experiment will need in order to say with any confidence that the program has an impact of a certain size. This is an important part of experimental design, and here we walk through the factors considered.

How Small an Impact Do We Need?

The size of the sample needed depends on how small an effect we need to detect. Experiments require a larger sample to detect a smaller impact. It is very important to make an educated guess as to the range of impact a program like the one being tested typically has. From a practical point of view, it is also important to know the smallest potential impact that would be considered educationally useful. As a hypothetical example, using percentile ranks as the measure of impact, we may predict that an intervention of this type can often move an average student 15 percentile points. As a sensible matter for educators, however, an improvement as small as 10 percentile points may have practical value. The researcher may then set the smallest effect of interest to be 10 points—the intervention may do better but if it makes less than a 10 point difference, the practical value will be no different than zero. We can call this the "minimum required effect size." It is necessary to decide on this value as part of the power analysis, since the number of units needed in the sample is related to how small an effect we need to detect. Conversely, with a particular number of units available, we want to know how small an effect we can detect—the so-called "minimum detectable effect size" (MDES). In some cases, there may be positive effects that we can't detect because they are lower than the MDES.

How Much Variation is there Between Teachers?

When we randomize at the teacher level, but the outcome of the interest is a test score of students associated with those teachers, we pay special attention to the differences among teachers. The greater the differences among those units, the more units we need in the experiment to detect the impact of the intervention. This is because the variation among teachers that isn't due to the program adds noise to our measurement, which makes the effect of the intervention—the signal—harder to detect. A larger sample effectively allows us to reduce the level of the noise. If the differences among teachers are large and/or the differences within them are small, then the sample size that matters the most for the experiment is the number of teachers. If the differences among teachers are small so that most of the variation is attributable to differences among students within them, then the sample size that matters most is the number of students. A summary statistic that tells us how the variation is divided up among levels of analysis is the intraclass correlation (ICC). Technically it is the ratio of the variation in the outcome among teachers to the total variation. We assume that this is computed before the intervention. For this experiment we assumed a fairly conservative intraclass correlation of .20.

Randomization by Pairs

There are various ways to randomize teachers to experimental conditions. For this study, we use a matched-pairs design, where we identify pairs of similar teachers. First, we consider what the critical characteristics of teachers are that we believe affect performance. We use this information to pair teachers and then we randomize the members in each pair to the two conditions. Technically, this is a form of blocking and it usually increases the degree of certainty we have in the difference in the posttest scores that we measure between the program and control groups. Randomization took place in the first year of the experiment and the same groups were carried over into this phase. At that time, matches were based on the subjects teachers expected to teach for the coming year, class scheduling, and years of teaching experience. Twenty-two pairs of teachers were assigned using a coin toss to either the program condition or to control to ensure a balanced distribution.

How Much Value Do We Get From a Pretest?

In order to gain additional precision, we make use of other variables that we know will impact performance. In our experiments, a student's score on a pretest (which may be a test in a subject that is closely related to the outcome measure rather than the same test but given earlier) is almost always the variable most closely associated with the outcome. In this case, the pretest is a "covariate." By including the covariate we can increase precision by "removing" this source of variation in the results. Technically, a covariate-adjusted analysis is called an analysis of covariance (or ANCOVA). In almost all of our analyses we adjust for the effect of the pretest, which is a strong predictor of posttest performance. In planning this experiment, we assumed a fairly substantial correlation between the pre- and posttests (.80[1]). In a power analysis determining the number of teachers we will need, a good pretest correlation will increase precision and thereby require fewer teachers to detect the same level of impact.

Are There Subgroups of Particular Interest?

Often we are interested in whether a program has more of an impact for a particular subgroup than for others or for some teachers but not others. Where the subgroup is identified within each randomized unit, that is, where each randomized unit has some portion of that subgroup, the impact on the power analysis is usually minimal. However, if our subgroup of interest is a subtype of the unit of randomization, then in most cases, we will need to include more units in the experiment in order to have enough units of each type. In the current experiment, we are interested first of all in separating the results for Algebra and Geometry students. This effectively split our teacher and student sample in half and, for most of our estimates has a major impact on our power analysis. The other characteristics of interest—differing levels of the pretest, gender, and English proficiency—are at the student level and so are divisions within our units of randomization.

How Confident do we want to be in the Results?

We described uncertainty in terms of the likelihood that if we ran the experiment again with a different sample from the same district we would get the same result. Although the results would never be exactly the same, we can design the experiment so that the different results that we

[1] That is, we assume that .80*.80=.64 is the proportion of variance in the outcome (i.e., the R-squared) that is accounted for by the covariate, in either condition.

would get would likely fall within a certain range. An experiment that produces a very high level of confidence that the results would be very similar requires a larger number of units than an experiment what produces a lower level of confidence or a wider range of likely outcomes for the other hypothetical experiments. The final step in the power analysis is to decide an acceptable or tolerable level of uncertainty. Conventionally, researchers have called for a high level of certainty, specifically, that getting a result like that observed would happen only 5% of the time if the program didn't actually have an impact. For the purpose of the power analysis for this experiment, we used that criterion although, as we explain later, we report the results using a range of confidence levels up to 20%.

Sample Size Calculation for This Experiment

Taking all the above factors into consideration, and with the number of teachers that were available for this study, we estimated that the smallest effect size that we can detect is a difference of 13 percentile points for Algebra and 15 for Geometry. This means that if the 50th percentile Algebra student in the program group were placed at the end of the experiment in control he or she would be that many points higher (or lower) than the 50th percentile student in the control group. As we explain later in this section, we can also express these as standardized effect sizes or portions of a standard deviation. In that metric the MDES for Algebra is 0.34 and for Geometry 0.38. This power analysis takes into account that we are treating the Algebra and Geometry teachers essentially as two separate experiments.

The TI-Navigator System

The experimental program consists of the TI-Navigator system with graphing calculators as compared to classrooms with the graphing calculators alone. Separate professional development training was provided for both sets of technology, the TI-Navigator system with graphing calculators and the graphing calculator alone. The technologies were intended to be integrated with the school's standard Algebra I and Geometry curricula. In ESUHSD, teachers primarily use the Prentice Hall series and the McDougal Littell series to teach Algebra I and Geometry. In SDUSD, teachers primarily use the McDougal Littell series or the Key Curriculum Press (*Discovering Algebra* or *Discovering Geometry)* to teach Algebra and Geometry.

Training/Professional Development

The *GC+Nav* group received a three-day training on using the TI-Navigator. The *GC+Nav* training was spread over one week and allowed for the *GC+Nav (A)* and *GC+Nav (B)* groups to cover separate topics. The *GC+Nav (A)* group met at the beginning of the week and further developed the TI-Navigator skills introduced in the January 2006. The *GC+Nav (B)* group met for the latter part of the week and were introduced to the basic functions and capabilities of the TI-Navigator. The two groups met together for one day in the middle of the week.

GC+Nav Materials

The TI-Navigator™ 3.0 system was deployed in August 2006 to all treatment teachers (Phase I *GC* teachers) who then became the *GC+Nav* treatment group for Phase II.

The TI-Navigator™ 3.0 system is designed to work with the TI graphing calculators and adds the following capabilities:

- Wireless communication between students' graphing calculators and the teacher's PC computer
- Activity center, quick polling, and screen capture activities

The *GC+Nav* materials also included the graphing calculators as described below for the control materials.

Control Materials

The graphing calculator was deployed in August 2006 to all control teachers and they remained the control group for Phase II.

The graphing calculator system has features that can be used in Algebra I and Geometry classrooms:

- Connectivity with a variety of presentation tools that allow opportunities for demonstrations of graphing and analysis techniques, data collection and analysis, and problem solving methods
- Applications for Algebra I including linear equations, functions, and inequalities
- Cabri® Jr. Dynamic Geometry – an interactive Geometry application for constructing, exploring, and analyzing of a variety of geometric objects
- Applications in probability theory and statistics
- Data acquisition through the use of sensors (CBL/CBR)
- Ability for students to share data and graphical displays via the presentation connector

TI also provided all study teachers a standard notebook computer with TI-SmartView software (a calculator emulator), a data projector, and calculator-based ranger (CBR) units. The text materials provided by TI for the trainings are discussed in the Implementation section, as are the workshops attended by the teachers.

Implementation Schedule

Materials were deployed and training was provided as summarized in Table 1.

Table 1. Research Milestones

Milestone	Date
Navigator and graphing calculator training workshop at ESUHSD and SDUSD	August 14 – 18, 2006
Teacher surveys begin	October 6, 2006
NWEA pretest administration	September – October 2006
Classroom observations at SDUSD	October 23 – 27, 2006
Classroom observations at ESUHSD	October 30 – November 3, 2006
Classroom observations at SDUSD	January 22 – 24, 2007
Follow-on training for *GC+Nav (B)* at ESUHSD	February 16 – 17, 2007
Classroom observations at SDUSD	February 28 – March 2, 2007
Classroom observations at ESUHSD	March 26 – 30, 2007
CST posttest	April – May 2007
Teacher interviews	May 2007
Classroom observations at ESUHSD and SDUSD	May 2007
Teacher surveys conclude	May 15, 2007
NWEA posttest	May 7 – June 8, 2007
Debrief meetings	June 2007

Site Descriptions

East Side Union High School District

East Side Union High School District (ESUHSD) is located in San Jose, California. San Jose is a large city located approximately 50 miles south of San Francisco. The total population of San Jose is 894,943 (U.S. Census, 2000). ESUHSD spends approximately $7,688 per student (NCES, 2005). ESUHSD has 21 schools serving grades 9, 10, 11, and 12, one school serving grades K-12, and one alternative school serving the community. Table 2 provides information about the entire district, including the high schools that participated in the study.

Table 2. Demographics of the East Side Union High School District

East Side Union High School District	
Total Schools	23
Total Teachers	1,148.5
Student to Teacher Ratio	22.2
Grade structure (i.e., K-6)	9-12[a]
Student Population	25,496
Migrant Students	911
ELL Students	7,012
White	13%
Black	5%
Hispanic	44%
Asian	27%
Pacific Islander	1%
Filipino	10%
American Indian/Native Alaskan	0%
Multi Racial	0%

[a] One alternative school in the district, not in the experiment, includes K-12

Source: CCD Public school data 2004-2005 school year.

The research was conducted at nine schools sites in East Side Union High School District serving grades 9, 10, 11, and 12. The specific demographics for each of the schools are reported in Table 3.

Table 3. Participating High Schools' Demographics

Participating East Side Union High Schools[a]									
School ID	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
Student Population	2,744	2,090	1,123	4,005	1,563	1,509	2,006	2,124	2,181
Grades	9-12	9-12	9-12	9-12	9-12	9-12	9-12	9-12	9-12
Scheduling[b]	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Block	Block
Title I	No	No	Yes	Yes	Yes	Yes	No	No	No
Free/Reduced	28%	15%	35%	29%	50%	57%	30%	12%	13%
White	24%	17%	10%	7%	2%	2%	10%	16%	50%
Black	9%	5%	2%	4%	2%	3%	6%	5%	5%
Hispanic	38%	19%	74%	34%	57%	74%	54%	23%	25%
Asian	23%	48%	7%	34%	30%	10%	18%	44%	15%
Pacific Islander	1%	0%	1%	1%	1%	2%	1%	1%	1%
Filipino	4%	10%	6%	19%	7%	9%	12%	12%	3%
American Indian/ Native Alaskan	1%	0%	1%	0%	0%	0%	0%	0%	1%
Multi Racial	0%	0%	0%	0%	0%	0%	0%	0%	1%

[a] CCD Public school data 2004-2005 school year

[b] East Side Union High School District, 2006

San Diego Unified Schools

San Diego Unified Schools (SDUSD) is located in San Diego, California. San Diego is a large city located approximate 130 miles south of Los Angeles. The total population of San Diego is 1,223,400 (U.S. Census, 2000). SDUSD spends approximately $8,482 per student (NCES, 2005). SDUSD has 221 schools serving grades K-12. Table 4 provides information about the entire district, including the high schools that participated in the study.

Table 4. Demographics of the San Diego Unified School District

San Diego Unified School District	
Total schools	221
Total teachers	7,189.7
Student to teacher ratio	18.7
Grade structure (i.e. K-6)	K – 12
Student population	134,709
Migrant students	142
ELL students	37,076
White	26%
Black	14%
Hispanic	43%
Asian	9%
Pacific Islander	1%
Filipino	7%
American Indian/Native Alaskan	1%
Multi racial	0%

Source: CCD Public school data 2004-2005 school year.

The research was conducted at 11 schools sites in SDUSD serving grades 6 through 12. The specific demographics for each of the schools are reported in Table 5.

Table 5. Participating High Schools' Demographics

Participating San Diego Unified Schools[a]											
School ID	**11**	**12**	**13[b]**	**14**	**15**	**16**	**18**	**20**	**21**	**22**	**35[b]**
Student population	2920	434	471	465	1462	2001	2568	491	1036	2474	462
Grades	9-12	9-12	9-12	7-12	9-12	9-12	9-12	9-12	6-8	9-12	9-12
Scheduling[c]	Daily	Block	Block	Block	Daily	Daily	Daily	Block	Daily	Daily	Block
Title I	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
Free/ Reduced	51%	81%	65%	69%	42%	45%	32%	75%	37%	33%	78%
White	5%	3%	9%	13%	44%	32%	25%	6%	47%	51%	2%
Black	20%	30%	15%	17%	6%	17%	10%	16%	13%	11%	4%
Hispanic	28%	50%	74%	64%	44%	34%	13%	77%	27%	28%	94%
Asian	4%	17%	1%	2%	4%	11%	19%	0%	8%	8%	1%
Pacific Islander	3%	1%	n/a	1%	1%	2%	2%	0%	1%	1%	n/a
Filipino	40%	0%	n/a	2%	1%	5%	30%	1%	3%	1%	n/a
American Indian/ Native Alaskan	0%	0%	1%	1%	1%	1%	0%	0%	1%	1%	0%
Multi-Racial	0%	0%	n/a	0%	0%	0%	0%	0%	0%	0%	n/a

[a] CCD Public school data 2004-2005 school year

[b] Asian ethnic category includes Pacific Islander and Filipino

[c] San Diego Unified School District, 2006

Data Sources and Collection

The data for this study consist of student outcomes, demographics, and classroom observations. In addition to conducting formal and informal teacher interviews, we also collected 15 web-based survey responses from all participating teachers in each group. We retrieved TI-Navigator system log files from the *GC+Nav* classroom computers to confirm Navigator use.

District Supplied Information

The data requested from the school districts included records for the students who were taught by participating teachers as well as other background information. Specifically, the districts were asked to provide the following data:

- Student name or unique ID
- Gender
- National School Lunch Program status (proxy for socio-economic level)
- Ethnicity
- Home language
- English learner status
- Disability status (whether or not student is has a disability or is in Special Education but not the specific condition)
- Age
- Classroom teacher
- School the student attends

One site was unable to provide National School Lunch Program status.

All student and teacher data having any individually identifying characteristics were stripped of such identifiers, and the data were stored using security procedures consistent with the provisions of the Family Educational Rights and Privacy Act (FERPA).

Achievement Measures

For the study, we used two outcome measures: the California Standards Test (CST) and the Northwest Evaluation Association (NWEA) End of Course tests.

As our pretest covariate, we used one pretest that all students in the study took: the California aligned NWEA 6+ General Math Surveys. Students came to Algebra I and Geometry from a variety math classes and took different CST math tests the year before. Since the different CST math tests tap different constructs and are not comparable, scores from the previous year CST could not be used as a pretest. In all cases, the pretest covariate was the NWEA pretest.

During the testing windows, multiple versions of the pretest and posttest had to be administered to accommodate field conditions. Statistical comparisons showed different versions of each test had different score distribution patterns that suppressed or inflated scores in relation to each other. As we explain in more detail below, to make the different versions of tests comparable, we transformed the scores so that they were on a common scale with the mean score set at zero and all other scores placed using a common metric in their relative position to the mean (called a z-transformation).

The Math CST Posttest

The CST is part of the Standardized Testing and Reporting (STAR) program through the Testing and Accountability office in the California Department of Education. The criterion referenced tests are given each year to all students in grades 2 through 11 and are typically administered by school faculty and staff. We did not observe the CST posttest administration.

The outcome measures we used were student test scores on the Algebra I CST or Geometry CST. Students took a particular CST depending on the course just completed. The students in our sample come from different grades but for either subject they all take the same test, so vertical alignment is not an issue. The two tests tap different constructs and cannot be compared. As a result, we treated Algebra I and Geometry as separate experiments.

Tests of skew and kurtosis revealed that both the Algebra and Geometry CST posttests were slightly skewed, but not sufficiently to warrant transforming them[2].

The NWEA Pretest and Posttest

The NWEA tests were used as a secondary assessment check against the CST. At the beginning of the study, ESUHSD already had several years of experience administering the NWEA as part of its district assessment practices. Using the NWEA created less of a burden on the ESUHSD as compared to introducing a different test. At SDUSD the NWEA was administered with support from our research team. We observed a sample of NWEA pretest and posttest administrations at each site.

The testing administration did not proceed smoothly at every school. There were computer network issues that delayed test administration and even precluded testing entirely on a given day. Combined with scheduling problems that prevented some classes from allotting more than a single class period to test, recovery of NWEA data was proved problematic and resulted in students missing a pretest or posttest score.

The NWEA is an assessment tool used by a large number of school districts nationwide for formative assessment. The NWEA can be administered in two formats, the MAP computer-adaptive test and the ALT paper-pencil test. Both tests are scored on a Rasch unIT (RIT) scale, a measurement scale developed to simplify the interpretation of test scores. This scale is used to measure student achievement and student growth on an equal-interval scale so that a change of one unit indicates the same change in growth, regardless of the actual numerical values. RIT scores range from about 150 to 300 and indicate a student’s current achievement level along a curriculum scale for a particular subject. In addition to the overall RIT score, the measures also provide goal score ranges to aid in the identification of a student's instructional levels.

The NWEA MAP Tests

The NWEA MAP used is a state-aligned, computer-adaptive assessment program that provides educators with information designed to improve teaching and learning. This adaptive test reflects the instructional level of each student and measures growth over time. The test goal structures are created through an alignment process that links state standards documents to the NWEA item bank.

[2] The plot of both CST Geometry and CST Algebra posttests against the NWEA pretest exhibited a slight curvilinear trend. We ran all analyses both with and without the addition of a quadratic term to account for the non-linearity. Usually the addition of this term did not change the result of interest. We report results for the simpler model without the quadratic term. We indicate whenever addition of the quadratic term led to a difference in the result.

ESUHSD used the NWEA MAP test during Phase I of this study as part of its standard assessment procedures. At the beginning of Phase II, the district discontinued the NWEA MAP, but agreed to administer the NWEA MAP to students in the study classes. Since ESUHSD was already familiar with using the NWEA, we provided minimal support with the organizational and technical aspects of the test at specific schools. The NWEA was conducted by ESUHSD with assistant principals coordinating the testing at individual school sites and school IT staff and classroom teachers proctoring their own classes.

At the beginning of the first year of the study, SDUSD was not familiar with the NWEA tests and lacked the computer infrastructure to administer the MAP. Midway through year I, SDUSD agreed to pilot the MAP test at two schools. For the second-year study reported here, the MAP test expanded to three schools with support by the research staff. School administrators, study teachers, and IT staff lead the organizational, logistical, and local technical aspects of the testing, while we created and provided the testing program and assisted with proctoring.

The NWEA ALT Tests

The NWEA ALT is a state-aligned, leveled, paper-pencil assessment that provides educators with information designed to improve teaching and learning. The ALT tests were administered exclusively at SDUSD because they did not have the computer infrastructure or IT personnel to administer the MAP test district wide. We were assured by the NWEA testing company that the ALT and MAP versions of the test are comparable; however, upon closer examination, scores from the two formats displayed different distribution patterns. While the overall RIT scores for each subject test are supposed to be comparable, statistical analysis showed the MAP and ALT test scores having different distribution patterns within each subject test.

The NWEA End of Course Geometry MAP and ALT

The NWEA Geometry End of Course Test has goal structures aligned to California's geometry standards and the topics tested are specifically associated with geometry content. In the Geometry End of Course Test, the goal structures are Spatial Relationships, Measurement, Geometric Relationships, and Problem Solving.

The NWEA End of Course Algebra I MAP and ALT

The NWEA Algebra I End of Course Test has goal structures aligned to California's Algebra I standards and the topics tested are specifically associated with Algebra I content. In the Algebra I End of Course Test, the goal structures are Linear Equations, Quadratic Equations, Algebraic Operations, and Problem Solving.

The NWEA General Math Goals Survey CA 6+ ALT

The NWEA General Math 6+ ALT was used as the pretest for one school district. The Math Goals Survey ALT is a two-step process where a student initially takes a short locator test that indicates that student's math ability and determines which leveled test booklet is appropriate for the student's current achievement level. The student is then administered the appropriate level test booklet to complete the test.

The NWEA General Math Goals Survey 6+ MAP V1 and V2

The NWEA General Math Goals MAP test had two versions administered, V1 and V2. The overall RIT score of the two versions are said to be comparable, but have different strands aligned to the California state standards. The general goal areas or strands within a state's standards become the goals measured. In the Math Goals Survey 6+ CA V2, Number Sense, Algebra and Functions, Measurement and Geometry, Statistics and Probability, and Mathematical Reasoning are the strands aligned to California's standards. In the Math Goals Survey 6+ CA V1, the strands are Number Sense, Estimation Computation, Algebra and Functions, Geometry, Measurement, and Statistics, Data, and Probability.

Statistical examination showed the RIT scores of the V1 and V2 to have different score distribution patterns.

The NWEA General Math Survey 6+ MAP V1 and V2

In ESUHSD, fewer than 30 students mistakenly took the incorrect NWEA pretest. The General Math Survey 6+ pulls from the same test item bank, but is about half as long as the General Math Goals Survey 6+. Since the General Math Survey has fewer items, it does not have the precision to break the score into goal structures. The General Math Survey 6+ has two versions of the test, V1 and V2. Fewer than ten students took V1 and fewer than 20 students took V2.

Scale Transformation of NWEA Scores

Pretests. Altogether each student took one of the four versions of the NWEA pretests that are described above. Most of the students took the Math Goals Survey 6+ V1. To adjust for the lack of observed comparability among the pretests, scores were transformed within each version so that they were expressed in terms of standard deviations away from the mean (i.e., they were z-transformed). These transformations were performed with Algebra and Geometry students combined (the pretests measured general math ability and so were not specifically aligned with either subject). Cases were then separated by subject (Algebra or Geometry) and pretests were centered on their respective means to allow an easier interpretation of the results.

Posttests. Algebra and Geometry outcomes are reported separately. As described above, the NWEA test can be either a paper-and-pencil version (ALT) or a computerized-adaptive version (MAP). A large majority of students took the MAP. The two versions are not equated so we z-transformed NWEA outcomes within each of the two types of scale, separately for each subject. (We report results for the analyses of ALT and MAP outcomes combined, but we also ran each analysis for MAP outcomes only. We note whenever there is a discrepancy between the results.)

Testing Schedule

The pretest and posttest for this study were administered by the school districts. School administrators coordinated the logistics of testing and school IT staff supported the technical aspects associated with the MAP. School faculty and staff proctored their own classes during each administration during the testing window. We also provided both districts with some technical and logistical support during the test administration. We observed a sample of NWEA pretest and posttest administrations. The testing proceeded as planned with no major problems.

Schools administered the NWEA pretest between September and October 2006. One site had experience administering the NWEA test as part of its standard assessment procedures. The district was familiar with the MAP procedure and conducted the test administration themselves. The other site had experience administering the NWEA as part of the study from Phase I.

One school site ran on a 4x4 schedule. Under this schedule, the year-long math course is compacted into a one-semester math course with longer daily periods. For this site, the school administered NWEA posttests at the end of the first semester in January 2007. NWEA pretests were immediately administered in February 2007 at the beginning of the second semester to a new set of students.

The schools administered the CST posttest in April 2006 and the NWEA posttest in May and June 2006. The months of April, May, and June are typically high assessment administration months in many school districts. Teachers noted that some students suffered from testing fatigue and attempted to space their tests. We did not observe CST posttest administration.

Observational and Interview Data

In addition to quantitative data, we also collected qualitative data over the entire period of the experiment, beginning with the randomization meeting and ending with the academic calendar of the district in June 2007. Training observations, classroom observations, informal and formal interviews, multiple teacher surveys, email exchanges, and phone conversations are used to provide both descriptive and quantitative evidence of the implementation.

In general, observational data are used to inform the description of the learning environment, instructional strategies employed by the teachers, and student engagement. These data are minimally coded. For the observation protocol used, please see Appendix D1 and D2 of the Phase I Final Report.

Interview data are used to elaborate survey responses, characterize the teacher's schedule, and to provide descriptions of the overall experience teaching with the *GC+Nav* and control materials.

Survey Data

The quantitative survey data are reported using descriptive statistics; these are summarized by individual teacher and by assignment group (program and control) and are compared by group. The free-response portions of the surveys are minimally coded. Survey data are used to quantify the extent of exposure to the materials (opportunities to learn with the curriculum).

Surveys were deployed to both *GC+Nav* and control group teachers beginning on October 6, 2006 and continued on a bi-weekly basis until May 18, 2007. We calculated response rates as simple percentages based on the ratio of actual received responses to the number of expected responses. There were 16 teachers in the *GC+Nav* group and 19 teachers in the control group. A total of 525 individual surveys were deployed with an overall response rate of 99.4% for both groups, a 98.7% response rate for the *GC+Nav* teachers and a 100% response rate for the control teachers.

The survey topics were developed to account for the various aspects of teacher and student actions associated with instruction and learning. For example, in order to characterize the average time teachers and students spent using the TI-Navigator and graphing calculators, we used a repeated question strategy. On surveys 1 through 13, we asked teachers to estimate the instructional time they spent teaching with the technology and time students spent using the technology. We received a 100% response rate. These questions, together with questions regarding the types of activities, allow us to draw inferences about how time was devoted to using the TI technology in both the *GC+Nav* and control groups.

Formation of the Experimental Groups

This section describes the sample that we use to determine the impact of TI-Navigator. The sample consists of two randomly assigned groups of teachers. We describe this sample as being formed initially through the random assignment but modified somewhat through attrition or loss of teachers and students at different points during the experiment for a variety of reasons. Ideally, by randomizing assignment into the two conditions, the groups should look the same in terms of important characteristics such as demographic composition, achievement, and teacher characteristics. In addition because we paired teachers, we can expect somewhat better balance than we would have if we hadn't first balanced them on these characteristics. However, by chance (as well as the imprecision of the pairing) the groups are never exactly balanced and may indeed be unbalanced on important characteristics likely to affect the outcome. Moreover, the loss of teachers and students during the experiment may introduce a bias if, for example, treatment or *GC+Nav* teachers are more likely to drop out of the program than control because of the extra burden. Therefore in this section we inspect the teachers and students and look in particular at the balance between the *GC+Nav* and control groups. We examine whether there was differential attrition between the *GC+Nav* and control groups both overall and with respect to subgroups of students and teachers. We also inspect the final sample that is available for determining impact and check whether the *GC+Nav* and control groups are balanced on important characteristics, recognizing that imbalance may have entered into the sample both because of "unlucky" randomization and through a biased attrition process. For this accounting, we focus on the data available for NWEA and CST results.

The tables that follow provide an account of the teachers and students in the experiment showing attrition and the resulting sample used the calculate impact. Two tables show the sample used for the

Algebra calculations, one for the sample used with the NWEA results and the other for the sample used for the CST results. Two similar tables show the Geometry sample. The first line of each table shows the number of teachers starting Phase II of the experiment. This includes teachers who completed the first year, teachers who were counted as attrition for various reasons but, nevertheless, available for Phase II, and two teachers who joined the study, one being randomized to the *GC+Nav* group and the other to control. Seven of the teachers had assignments in both subjects so the numbers on the top line across the two topics is greater than the number of teachers. Also, assignments to Algebra and Geometry are not necessarily the same from one year to the next. Finally, some of what shows as attrition in the table is actually associated with classes assigned to the teacher but in which the technology was not used. So for example, a teacher, who in Phase I was assigned to Algebra but in Phase II assigned both Algebra and Geometry classes, may have used the technology only in the Algebra classes. For completeness, this teacher (and the students in the Geometry classes not used in the study) will appear as attrition from the Geometry table. While this teacher choice has the potential for bias, we find that in each case, the choice appears unrelated to the program under study.

Algebra

Attrition

Table 6 traces the sample of Algebra students and their teacher for whom we have NWEA outcome data. The table shows there was attrition of one control teacher from the group of randomized teachers. The reason was the burden of participation in surveys and testing. We received class rosters for 12 *GC+Nav* and 14 control teachers. However, four of these teachers did not include their Algebra classes in the study because they had used the technology in Geometry the previous year. During the middle of the year, schedule changes caused one *GC+Nav* teacher not to teach any eligible classes and was lost from the study. Additionally, difficulties administering a test that was not part of the school's typical testing schedule resulted in absent students and another substantial group of students to be lost prior to the posttest.

Table 6. Numbers of Units in the Algebra Experimental Groups and Attrition Over Time (Associated with NWEA Posttest)

	Control		*GC+Nav*	
Event	**No. of teachers**	**No. of students**	**No. of teachers**	**No. of students**
Randomization	15	n/a	12	n/a
(Loss prior to rosters)	(1)	n/a	(0)	n/a
Fall rosters received	14	747[a]	12	504[a]
(Loss before/at pretest)	(1)	(157)	(3)	(159)
Pretest scores received	13	590	9	345
(Loss before/at posttest)	(0)	(166)	(1)	(116)
Final count of units with pretest and posttest	**13**	**424**	**8**	**229**

[a] This sample excludes students with disabilities.

Table 7 traces the same sample of Algebra students and their teachers, but now examining those for whom we have CST outcome data. The table shows the same attrition pattern up to the posttest. Fewer students were lost at the posttest because CST administration is a state mandated test; however, a substantial number of students were lost from the control group. As a result, a similar set was available for examination of impact CST.

Table 7. Numbers of Units in the Algebra Experimental Groups and Attrition Over Time (Associated with CST Posttest)

Event	Control		*GC+Nav*	
	No. of teachers	No. of students	No. of teachers	No. of students
Randomization	15	n/a	12	n/a
(Loss prior to rosters)	(1)	n/a	(0)	n/a
Fall rosters received	14	747[a]	12	504[a]
(Loss before/at pretest)	(1)	(157)	(3)	(159)
Pretest scores received	13	590	9	345
(Loss before/at posttest)	(0)	(137)	(1)	(66)
Final count of units with pretest and posttest	**13**	**453**	**8**	**279**

[a] This sample excludes students with disabilities.

Characteristics of the Initial Sample

In Table 8 we compare the composition of the control and *GC+Nav* groups at the point we received the rosters. For each of the characteristics of this sample, we conducted a statistical test[3] to determine the likelihood of obtaining a chance imbalance as large as, or larger than, the one observed. Of course the randomization assures us that any imbalance was a result of chance, and is not an indication of selection bias, but it is useful to examine the actual groups as formed to see whether the amount of imbalance is something we would expect to see less than 5% of the time. We observe that the *GC+Nav* group had a higher portion of English proficient students and that the group had a higher average pretest score. Ethnicities were also not balanced; however, we did not expect that such differences would make a difference beyond potential association with achievement levels, which are addressed in our estimation of the impact.

[3] For the categorical variables, we used the Fisher Exact Test. For the continuous variables, we used a t test. In both cases the criterion for significance was set at <.05.

Table 8. Characteristics of Algebra Study Sample

	Control group	*GC+Nav* group	Less than 5% chance of seeing this much imbalance
Student characteristics			
English proficient[a]	484 (66.2%)	362 (71.8%)	Yes
Male[b]	409 (52.3%)	300 (53.2%)	No
White[c]	79 (10.8%)	70 (13.9%)	
Black[c]	54 (7.4%)	53 (10.5%)	
Hispanic[c]	490 (66.8%)	300 (59.6%)	Yes
Asian[c]	107 (14.6%)	79 (15.7%)	
Native American[c]	3 (0.4%)	1 (0.2%)	
Mean pretest score	-0.27	-0.17	Yes
Teacher characteristics			
Less than 4 years teaching experience	2 (15.4%)	2 (25.0%)	No

[a] Information about English proficiency status is missing for 16 students.

[b] Information about gender includes counts of students with disabilities.

[c] Information about ethnic background is missing for 15 students; balance was checked simultaneously for all ethnic groups.

Geometry

Attrition

As shown in Table 9, Table 10, and Table 11, we accounted for the Geometry group in the same way. Before the school year began, one control teacher retired from teaching. After rosters were received, we determined that five teachers were not available for the Geometry *GC+Nav* group. Two of these had schedule changes and no longer taught an eligible class. Two teachers taught both subjects but only used the technology in Algebra because that was the subject in which the technology was used last year. One teacher left the study because of the burden of surveys and testing. Three teachers were lost during the year. Two *GC+Nav* teachers and one control teacher left due to schedule changes.

Table 9. Numbers of Units in the Geometry Experimental Groups and Attrition Over Time (Associated with NWEA Geometry Posttest)

	Control		*GC+Nav*	
Event	No. of teachers	No. of students	No. of teachers	No. of students
Randomization	9	n/a	14	n/a
(Loss prior to rosters)	(1)	n/a	(0)	n/a
Fall rosters received	8	303[a]	14	560[a]
(Loss before/at pretest)	(0)	(49)	(5)	(196)
Pretest scores received	8	254	9	364
(Loss before/at posttest)	(1)	(55)	(2)	(107)
Final count of units with pretest and posttest	**7**	**199**	**7**	**257**

[a] This sample excludes students with disabilities.

Table 10. Numbers of Units in the Geometry Experimental Groups and Attrition Over Time (Associated with CST Geometry Posttest)

	Control		*GC+Nav*	
Event	No. of teachers	No. of students	No. of teachers	No. of students
Randomization	10	n/a	16	n/a
(Loss prior to rosters)	(1)	n/a	(0)	n/a
Fall rosters received	8	303[a]	14	560[a]
(Loss before/at pretest)	(0)	(49)	(5)	(196)
Pretest scores received	8	254	9	364
(Loss before/at posttest)	(1)	(57)	(2)	(87)
Final count of units with pretest and posttest	**7**	**197**	**7**	**277**

[a] This sample excludes students with disabilities.

Characteristics of the Initial Sample

The initial sample of Geometry students and teachers were well balanced except with respect to ethnic makeup.

Table 11. Characteristics of Geometry Study Sample

	Control group	*GC+Nav* group	Less than 5% chance of seeing this much imbalance
Student characteristics			
English proficient[a]	202 (66.7%)	378 (70.0%)	No
Male[b]	144 (47.5%)	264 (47.4%)	No
White[c]	40 (13.2%)	88 (15.9%)	
Black[c]	25 (8.3%)	60 (10.8%)	
Hispanic[c]	136 (44.9%)	286 (51.6%)	Yes
Asian[c]	101 (33.3)	120 (21.7%)	
Native American[c]	1 (0.3%)	0 (0.0%)	
Mean pretest score	.38	.46	No
Teacher characteristics			
Less than 4 years teaching experience	1 (14.3%)	2 (28.6%)	No

[a] Information about English proficiency status is missing for 20 students.

[b] Information about gender is missing for 3 students but includes counts of students with disabilities.

[c] Information about ethnic background is missing for 6 students; balance was checked simultaneously for all ethnic groups.

Differential Student Attrition During the Experiment

Table 12 shows the mean pretest score for each set of students where the group is defined according to the posttest score used in the examination of impact. The question here is whether the students for whom we had pretests were differentially lost during the experiment. The sets of remaining students were defined by whether or not we had posttests for them. The table shows that, for all sets of students, the remaining students had significantly higher scores than the students who were lost.

Table 12. Mean Pretest Scores of Students Who Were Lost During the Experiment Compared to Those Remaining in the Sample

	Mean pretest of lost students	Mean pretest of remaining students	Significant potential for bias
Outcome measures			
NWEA Algebra	-0.41	-0.15	Yes
CST Algebra	-0.45	-0.18	Yes
NWEA Geometry	0.15	0.52	Yes
CST Geometry	-0.15	0.57	Yes

All scores have been z transformed, i.e., converted to standard deviation units.

As an explanation for this, teachers noted that there is a “point of no return” for students taking a course, wherein regardless of what the student does (more homework, extra credit, score all A’s on future exams) a failing grade will not improve, and the student will fail the class. When that point is reached, the low-performing student either stops attending class regularly or stops completely. Even when he or she is present in class, little or no work is done. Consequently, low-performing students are not in attendance toward the end of the semester or end of the school year when posttests typically are administered. The implication of this pattern is that the experimental results do not necessarily apply to the lowest scoring students in the districts.

Comparison of the Final Sample to the District Populations

Table 13 shows the key demographics and composition of the two districts and compares that to the Algebra and Geometry samples. We provide this for inspection but do not attempt to calculate the extent to which our sample is representative of the districts. The two districts are themselves different in size and, to some extent, in ethnic composition; for example, the Asian population in ESUHSD is larger.

Table 13. Comparison of District and Final Study Population

	Composition of school district populations		Composition of the study samples after attrition	
	ESUHSD population	SDUSD population	Algebra sample	Geometry sample
Total Schools	23	221	19	15
Total Teachers	1,148.50	7,189.7	23	15
Grades	9-12[a]	K-12	8-12	9-12
Student Population	25,496	134,709	1346	915
ELL Students	28%	28%	29%	29%
White	13%	26%	12%	15%
Black	5%	14%	9%	10%
Hispanic	44%	43%	64%	49%
Asian	27%	9%		
Pacific Islander	1%	1%	15%	26%
Filipino	9%	7%		
American Indian/Native Alaskan	0%	1%	0%	0%
Multi Racial	0%	0%	n/a	n/a

[a] One alternative school in the district, not in the experiment, includes K-12.

Adequacy of the Sample Size

The attrition experienced during this experiment occurred for a variety of reasons. In terms of our minimal detectable effect size (MDES), we find that it is higher than originally planned. For Algebra we will be able to detect a difference of 16% (or an effect size of 0.42). For Geometry we will be able to detect a difference of 20% (or an effect size of 0.53). The difficulty when MDESs are relatively large is that an effect smaller than that size could be present but not detected by the experiment. That is, we may conclude that there is no difference between the program and control when, in fact, there is a difference, but it is smaller than we can detect. We can rule out very large impacts but cannot confidently determine whether an effect smaller than the MDES may be attributable to the program.

Statistical Equations[4] and Reporting on the Impact of TI-Navigator

Setting Up the Statistical Equation

We use our data for students, teachers, and schools to obtain a system of statistical equations that allow us to obtain estimates of the direction and strength of relationships among factors of interest. The primary relationship of interest is the causal effect of the program on a measure of achievement. We use SAS PROC MIXED (from SAS Institute Inc.) as the primary software tool for these computations. The output of this process are estimates of effects as well as a measure of the level of confidence we can have that the estimates are close to the true parameter values.

Program Impact

A basic question for the experiment was whether, following implementation of the program, students in *GC+Nav* classrooms had higher algebra and geometry scores than those in control classrooms. Answering this is not as simple as comparing the averages of the two groups. The randomization gave us two groups that are equivalent to each other on average in every way, except that one receives TI-Navigator and the other one does not. But as we saw in the section on the formation of the experimental groups, in a single randomization we expect chance imbalances. Adjusting for these random differences gives us a more precise measure of the program's effect. It is also essential that we understand how much confidence we can have that there really is a difference between the two groups, given the size of the effect estimate that we obtain. To appropriately estimate this difference, our equation contains a term for TI-Navigator as well as terms for other important factors such as the student pretest score. The student's prior score is, of course, an important factor in estimating his or her outcome score. By including pretest as a term in the statistical equation, we are able to improve the precision of this estimate because it helps to explain a lot of the variance in the outcomes and makes it easier to isolate the program impact. We also have to account for the fact that students are clustered by classes and teachers. We expect outcomes for students who are in the same class or who have the same teacher to be dependent as a result of shared experiences. We have to add this dependency to our equation or else our confidence levels about the results will be artificially high.

Covariates and Moderators at the Student and Teacher Level

In addition to estimating the average impact, we also include in the equation other variables (called covariates) associated with characteristics of the students that we expect to make a difference in the outcomes for the students. For example, as was described above, we add the pretest score into almost all our statistical equations in order to increase precision. In addition, we consider whether there is a difference in the effect of the intervention for different levels of

---

[4] The term "statistical equation" refers to a probabilistic model where the outcome of interest is on the left-hand side of the equation and terms for systematic and random effects are on the right-hand side of the equation. The goal of estimation is to obtain estimates for the effects on the right-hand side. Each estimate has a level of uncertainty, which is expressed in terms of standard errors or *p* values. The estimate of main interest is for the treatment effect. In this experiment we model treatment as a fixed effect. With randomized control trials, the modeling equation for which we are estimating effects takes on a relatively simple form: Each observed outcome is expressed as a linear combination of a treatment indicator, one or more covariates that are used to increase the precision of the intervention effect estimate, and usually a series of fixed or random intercepts, which are increments in the outcome that are specific to units. As a result of randomization, the other covariates are distributed in the same way for both the treatment and control groups. For moderator analyses, we expand these basic models by including a term that multiplies the treatment indicator with the moderator variable. For dichotomous moderators, the coefficient for this term is the moderator effect of interest.

the covariates. For example, we consider whether the program is more effective for higher-performing students than for lower-performing students. We estimate this *difference* (between subgroups) *in the difference* (between the program and control groups) by including an interaction term in the statistical equation. This term multiplies together the variable that indicates whether the student is in the intervention group, and the covariate. We call covariates that are included in such analyses potential "moderators" because they may moderate—either increase or decrease—the effect of the program on student outcomes. The value of the coefficient for the interaction term is a measure of the moderating effect of the covariate on the effect of the program.

Fixed and Random Effects

The covariates in our equations measure two types of characteristics: 1) fixed characteristics that take on a finite set of values (e.g., there are only two levels of gender); or 2) a set of characteristics that is assumed to have a distribution over a population and where we treat the values that we measure as though they were a random sample from that larger population. The former are called "fixed effects," the latter, "random effects." Random effects add uncertainty to our estimates because they account for sampling variation, or the changes we would observe in the outcomes if we re-sampled units from the same population. Fixed effects produce less uncertainty but also limit the extent to which we can generalize our results.

We usually treat the units that were randomized as "random effects" so that in the statistical equations, our estimates reflect the degree of uncertainty that comes if we were to draw a different sample of units from the same population.[5] This allows us to argue for the generalizability of our findings from a sampling perspective. Treating the units that were randomized as fixed forces us to use other arguments if our goal is to generalize.

Using random or fixed effects for participating units serves a second function—it allows us to more accurately represent the dependencies among cases that are clustered together (e.g., students in classes). All the cases that belong to a cluster share an increment in the outcome—either positive or negative—that expresses the dependencies among them. An appropriate measure of uncertainty in our estimate of the program's effectiveness takes into consideration the relative variation in the outcome *within* and *between* higher-level units. All of our statistical equations include a student-level error term. The variation in this term reflects the differences we see among students that are not accounted for by all the fixed effects and other random effects in our statistical equation.

The choice of terms for each statistical equation is not rigid but depends on the context and the importance of the factors for the question being addressed. The tables reporting the estimates resulting from the computation will provide an explanation of these choices in table notes where necessary for technical review.

[5] Although we seldom randomly sample cases from a broader population, and in some situations we use the entire population of cases that is available, we believe that it is still correct to estimate sampling variation (i.e., model random effects). It is entirely conceivable that some part or the whole set of participants at a level end up being replaced by another group (for whatever reason) and it is fair to ask how much change in outcomes we can expect from this substitution.

Reporting the Results

When we run the computations on the data, we produce several results; among them are effect sizes, the estimates for fixed effects, and *p* values. These are found in all the tables where we report the results.

Effect sizes

We translate the difference between program and control groups into a standardized effect size by dividing the average group difference by the amount of variability in the outcome. The amount of variability is also called the “standard deviation” and can be thought of as the average distance of all the individual scores from the average score (more precisely, it is the square root of the average of squared distances). Dividing the difference by the standard deviation gives us a value in units of standard deviation rather than units of the scale used by the particular test. This standardized effect size allows us to compare the results we find with results from other studies that use different measurement scales. In studies involving student achievement, effect sizes as small as 0.1 (one tenth of a standard deviation) are sometimes found to be important educationally. When possible we also report the effect size of the difference after adjusting for pretest score and other fixed effects, since that adjustment provides a more precise estimate of the effect by compensating for chance differences in the average pretest of the program and control groups.

Estimates

We provide estimates to approximate the actual effect size. Any experiment is limited to the small sample of students, teachers, and schools that represent a larger population in a real-world (or hypothetical) setting. Essentially we are estimating the population value. When we report an estimate in a table, the value refers to the change in outcome for a one-unit increase in the associated variable. For example, since we code participation in the control group as 0 and participation in the *GC+Nav* group as 1, the estimate is essentially the average gain that we expect in going from the control to the *GC+Nav* group (while holding other variables constant).

P values

The *p* value is very important because it gives us a gauge of how confident we can be that the result we are seeing is not due simply to chance. Specifically, it tells us what the probability is that we would get a result with a value as large as—or larger than—the absolute value of the one observed when in fact there is no effect. Roughly speaking, it tells us the risk of concluding that the intervention has had an effect when in fact it hasn’t. This mistake is also known as a “false-positive” conclusion. Thus a *p* value of .1 gives us a 10% probability of drawing a false-positive conclusion. This is not to be confused with a common misconception about *p* values: that they tell us the probability of our result being true.

We can also think of the *p* value as the level of confidence, or the level of belief we have that the outcome we observe is not simply due to chance. While ultimately depending on the risk tolerance of the user of the research, we suggest the following guidelines for interpreting *p* values:

1. We have a high level of confidence when $p \leq .05$. (This is the level of confidence conventionally referred to as “statistical significance.”)
2. We have some confidence when $.05 < p \leq .15$.
3. We have limited confidence when $.15 < p \leq .20$.
4. We have no confidence when $p > .20$.

In reporting results with *p* values higher than conventional statistical significance, our goal is to inform the local decision-makers with useful information and provide other researchers with data points that can be synthesized into more general evidence.

Results

Implementation Results

In this section we characterize the implementation of TI-Navigator and the graphing calculators for this study. We refer the reader to the two interim reports for more detail on the composition of the experimental groups and technology trainings.

We triangulated teacher survey data, classroom observations, and teacher interview data collected during the course of the study to faithfully describe TI-Navigator and graphing calculator implementation.

The TI-Navigator logs data was used to confirm the usage time teachers reported on their surveys. The TI-Navigator logs recorded the date and time a teacher used the Navigator. By matching a log event to a specific survey, we were able to confirm the reliability of the teacher time self-reports. Anomalies in the log files, such as classes with a small number of students, prevented us from further use of TI-Navigator log files.

The TI-Navigator system did not have an explicitly defined set of implementation guidelines in terms of a TI-Navigator curriculum or number of activities to be used. Instead, expectations for TI-Navigator implementation were implicitly suggested by the skills and behaviors demonstrated during the TI-Navigator trainings that would be integrated with the teachers' existing math curricula. We concluded that teachers must have attended the trainings in order to understand how TI-Navigator was to be used. The trainings focused on using TI-Navigator to provide teachers with comprehensive and immediate feedback of student understanding. The trainings also highlighted TI-Navigator's ability to project and share student work in a cooperative workspace. To apply this in the classroom, teachers needed the resources to assemble the TI-Navigator system and to distribute graphing calculators to students. In addition, the frequency of TI-Navigator use would indicate how much exposure the students had to the system. Therefore, we measure implementation based on the following indicators:

- participating in training
- distribution and assembly of equipment and availability of resources
- the frequency of use
- the use of TI-Navigator as a formative assessment tool
- the use of TI-Navigator as a tool to create a cooperative learning workspace

We begin this section by discussing teacher experience with the equipment and teacher attrition. Next we describe the implementation data, which we organize by each of the indicators of implementation, comparing the TI-Navigator teachers to the graphing calculator teachers. Next we discuss barriers to TI-Navigator implementation. We conclude with a discussion of the differences between the *GC+Nav* and control conditions. Finally, we attempt to rate the level of implementation for the *GC+Nav* teachers.

Teacher Participants

At the start of the study, there were 19 *GC+Nav* teachers—nine from the *GC+Nav* A group (who had had experience with TI-Navigator in the second semester of the previous year) and 10 from the *GC+Nav* B group (newly introduced to TI-Navigator). One *GC+Nav* B teacher left the study due to school reassignment at the beginning of the school year and one *GC+Nav* B teacher did not participate due to the demands of the study. One *GC+Nav* A teacher left the school and was replaced by a substitute teacher who joined the *GC+Nav* B group. This substitute teacher ultimately left due to school reassignment, but completed the first semester of classroom surveys. One *GC+Nav* A teacher left the study due to health reasons after completing 10 classroom surveys. During the study one *GC+Nav* B teacher no longer taught eligible classes, but completed classroom surveys for the entire study. All but one of these teachers were excluded from the statistical analyses. All teachers who completed at least half of the classroom surveys are included in the following implementation section, where data were available.

The control group began with 19 teachers that remained throughout the study. The 19 teachers completed all classroom surveys and are included in the statistical analyses. One control teacher taught a complete Geometry class in the fall and a complete Algebra class in the spring and is included in the statistical analyses for both subjects. All 19 teachers are included in the following implementation section. The teacher who taught both subjects had implementation data for Algebra only.

Teacher Experience with TI-Navigator and Graphing Calculators

The *GC+Nav* A group had one semester experience using TI-Navigator during the previous year and the *GC+Nav* B group used TI-Navigator for the first time during the study. The entire *GC+Nav* group had an average of 6.8 years experience teaching with graphing calculators. The control group had an average of 6.4 years experience teaching with graphing calculators. One control teacher reported more than 10 years experience using graphing calculators and computers to teach mathematics.

Table 14. Number of Years of Experience Teaching with Graphing Calculators

	First year	2-5 years	6-9 years	10 or more
GC+Nav	1	9	2	4
Control	4	7	2	6

Indicators of Extent of Implementation

Training Participation

During the TI-Navigator trainings, the TI instructor modeled classroom usage of the TI-Navigator system. The TI instructor covered TI-Navigator assembly and configuration, TI-Navigator functions, and lesson structures using TI-Navigator. We videotaped TI-Navigator training sessions.

The *GC+Nav* A and *GC+Nav* B groups were each offered three days of initial TI-Navigator training. The *GC+Nav* A group was offered three days of follow-on training, but conflicting schedules forced the *GC+Nav* B group to have a two day follow-on training with an extra hour each day. For participants who missed part of the trainings, the trainer filled them in on what they had missed on the days they did attend. Participants who missed an entire training session were given DVD recordings of that session.

Table 15 shows the TI-Navigator training attendance. All members of the *GC+Nav* group participated in either the initial or follow-on training. One member of the *GC+Nav* A group missed the initial training and three members of the *GC+Nav* B group missed the follow-on training. School responsibilities, unavailability of substitute teachers, and competing priorities contributed to the low training attendance.

Table 15. TI-Navigator Training Attendance

	Total no. of teachers	Initial training: no. of days trained			Follow on training: no. of days trained				Avg. days participated
		0	2.5	3	0	1	2	3	
No. of *GC+Nav A* participants	8	1	0	7	0	0	1	7[b]	4.5
No. of *GC+Nav B* participants	8	0	1	7[a]	3	3	2	0	3.5

[a] One participant attended three half days of training and one participant missed two hours of training on two days.

[b] One participant attended two half days and one full day of training.

GC+Nav teachers were surveyed regarding how useful they found the training sessions on a scale of one to five (1=Poor, 2=Fair, 3=Good, 4=Very good, 5=Excellent). The responses are summarized in Table 16 and Table 17. Please refer to the Interim Report 2 for more detailed descriptions of the training.

Table 16. How Effective Was the TI-Navigator Training in the Following?

	Average rating	
Setting up the classroom on the computer	3.5	Good/Very good
Setting up the TI-Navigator hardware	3.1	Fair/Good

Table 17. Have You Used This TI-Navigator Feature and How Effective Was the Training of This Feature?

Topics	No. of teachers who have used feature[a]	Average rating	
Computer	14	4.14	Very good
Projector	14	3.88	Good/Very good
Taking Attendance	3	4.00	Very good
Screen Captures	10	4.00	Very good
Quick Polls	12	4.11	Very good
LearningCheck	6	2.75	Fair/good
Activity Center	7	2.63	Fair/good

[a]17 *GC+Nav* teachers responded.

The control group was offered five days of graphing calculator training. The graphing calculator training gave participants the opportunity to practice using the graphing calculator and associated hardware through a set of activities selected by trainers and researchers to match the control groups' Algebra and Geometry textbooks. Most of the control group participated, but five control group teachers did not attend any part of the training. Table 18 shows the control group training attendance.

Table 18. Days of Training Participation

		Number of training days participated				
	Total no. of teachers	0	3	4	5	Average days participated of those in attendance
No. of teachers	19	5	2	2	10[a]	4.6 days

[a]One teacher was absent three hours each day, two teachers missed 15-30 minutes of one day.

At the beginning of the school year, control teachers were surveyed regarding how useful they found the training session on a scale of one to five (1=Poor, 2=Fair, 3=Good, 4=Very good, 5=Excellent). The responses are summarized in Table 16, Table 17, and Table 19. Please refer to the Interim Report 2 for more detailed descriptions of the training.

Distribution of Graphing Calculators

TI-84 units were sent to each teacher to use for instruction. Findings from previous studies provide some evidence to indicate that when students have more access to calculators, both during class time and at other times, students score higher on end-of-course test scores (Heller, Curtis, Jaffe, and Verboncouer, 2005). TI-84 units were also sent for teachers to distribute to each student so the graphing calculator could be accessible at home. We collected survey data on the extent to which teachers used the TI-84 graphing calculator. All but one *GC+Nav* teacher reported teaching with the TI-84 and that their students used the TI-84. The teacher who reported not using the TI-84 used a scientific calculator and also reported that the students were using scientific calculators. A different teacher reported using the TI-84 to teach, but did not report any time using the TI-84 during the study.

Table 19. Have You Used This Feature and How Effective Was the Training of This Feature?

	No. of teachers who have used feature[a]	Average rating	
Graphing calculator			
Applications	6	3.78	Good/Very good
Lists	8	3.93	Good/Very good
CBRs	0	3.41	Good/Very good
Graphing	8	4.09	Very good
Study Card	1	2.84	Fair/good
Cabri Jr.	7	4.14	Very good
Study Card	1	3.00	Good
Hardware			
TI-SmartView	8	4.14	Very good
Projector	11	4.02	Very good
Computer	11	3.84	Good/Very good

[a]19 control teachers responded.

Control classrooms were supplied with enough graphing calculators for each student in an eligible class and asked to distribute the graphing calculators to students so they could be used at home and in class.

Table 20 and Table 21 show the student distribution of TI-84 graphing calculators to each student in the *GC+Nav* and control conditions.

Table 20. How Would You Characterize TI-84 Calculator Availability in Your lassrooms (*GC+Nav)*?

	Total no. of teachers	All students have been issued a calculator to take home and use in class	Some students check out calculators to take home	Students borrow from a class set and may NOT take home	No response
		#	#	#	#
12/1/2006[a]	17	8	6	6	1
2/2/2007	17	8	3	5	1

[a] Teachers were asked to check all that applied.

Table 21. How Would You Characterize Graphing Calculator Availability in Your Classrooms (control group)?

	Total no. of teachers	All students have been issued a calculator to take home and use in class	Some students check out calculators to take home	Students borrow from a class set and may NOT take home
		#	#	#
12/1/2006[a]	19	6	7	11
2/2/2007	19	8	2	9

[a] Teachers were asked to check all that applied.

Availability of Resources

We surveyed teachers in October and again in November as to whether they had the resources necessary to implement TI-Navigator, as displayed in Table 22. In October, three classrooms cited needing batteries and one classroom cited not being able to access the Internet.

At the beginning of the second semester in February, five teachers responded that they did not have the resources they needed. One teacher cited needing a TI-Navigator system that could easily be set up and taken down each day. One classroom needed replacements for stolen hardware, and one teacher cited needing help writing applications for TI-Navigator. Two teachers did not specify what was missing.

Table 22. Do You Have the Resources You Need to Implement the TI-Navigator Program?

Resources	No. of teachers	Yes	No
10/20/2006	17	13	4
2/16/2007	16	11	5

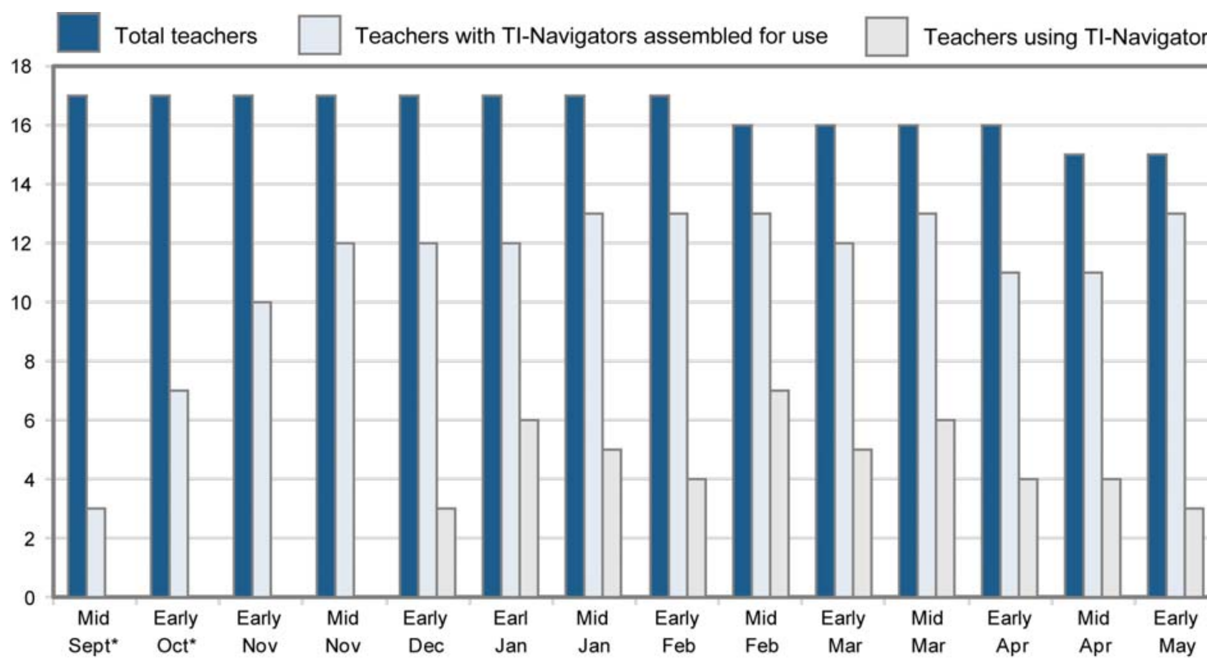
Although survey evidence suggests *GC+Nav* teachers had the resources needed to assemble TI-Navigator, the fact that some TI-Navigator systems were never assembled implies the something crucial was missing.

By contrast, all but one control teacher believed they had the resources to implement the graphing calculator. In October, one teacher reported needing the calculators to be engraved and in February, one teacher reported not knowing how to connect the TI-SmartView to the calculator.

Set-Up and Use

We collected observation and survey data that revealed when each TI-Navigator was assembled, as displayed in Figure 1. Starting in early December, we began collecting survey data regarding whether TI-Navigator was used in the classroom, also displayed in Figure 1.

Three *GC+Nav* A teachers assembled TI-Navigator at the start in September, but one of those teachers reported that the technology's fragility prevented TI-Navigator from being used regularly and eventually stopped using the system entirely. By early October, three additional *GC+Nav* A classrooms and one *GC+Nav* B classroom also assembled TI-Navigator. After early November, the number of teachers who indicated on their surveys that they had not assembled their TI-Navigator systems varied between two and five. Two *GC+Nav A* classrooms that did not assemble TI-Navigator the previous year did not assemble the system this year either. Of the 17 TI-Navigator systems deployed, a total of nine systems were recorded as having been used during the study and, when surveyed, at most seven systems were used at the same time.



*Data points come from observations.

Note. The total number of teachers decreases at February and April due to teacher attrition.

Figure 1. The Number of TI-Navigator Systems Assembled and Used

Frequency of Use

Time spent using TI-Navigator with graphing calculators provides an indication of the amount of classroom exposure to the TI-Navigator technology. Teachers were surveyed every other week regarding the amount of time spent using TI-Navigator. Teachers who had the TI-Navigator assembled spent approximately 18.4 minutes per week per class using TI-Navigator. When combined with time spent using graphing calculators, *GC+Nav* classrooms on average spent 40.9 minutes per week per period using the system.

Surveys and TI-Navigator log data of the nine implementing teachers suggest the technology was used continuously in one classroom. Seven classrooms that used TI-Navigator appeared to use it in short bursts for periods less than one week at a time, followed by several weeks of non-use. TI-Navigator computer logs were not recovered from one classroom that was observed to use TI-Navigator, but surveys suggest short bursts of technology use in that classroom as well.

The following graph plots the average time spent using TI-Navigator based on survey data over the course of the study. The graph loosely follows a cyclical pattern with the highest use appearing when surveys reported the majority of TI-Navigator systems being set up and at the beginning of the second semester. Usage also appears to drop as the winter holiday season began in December and the when state testing began in April and May.

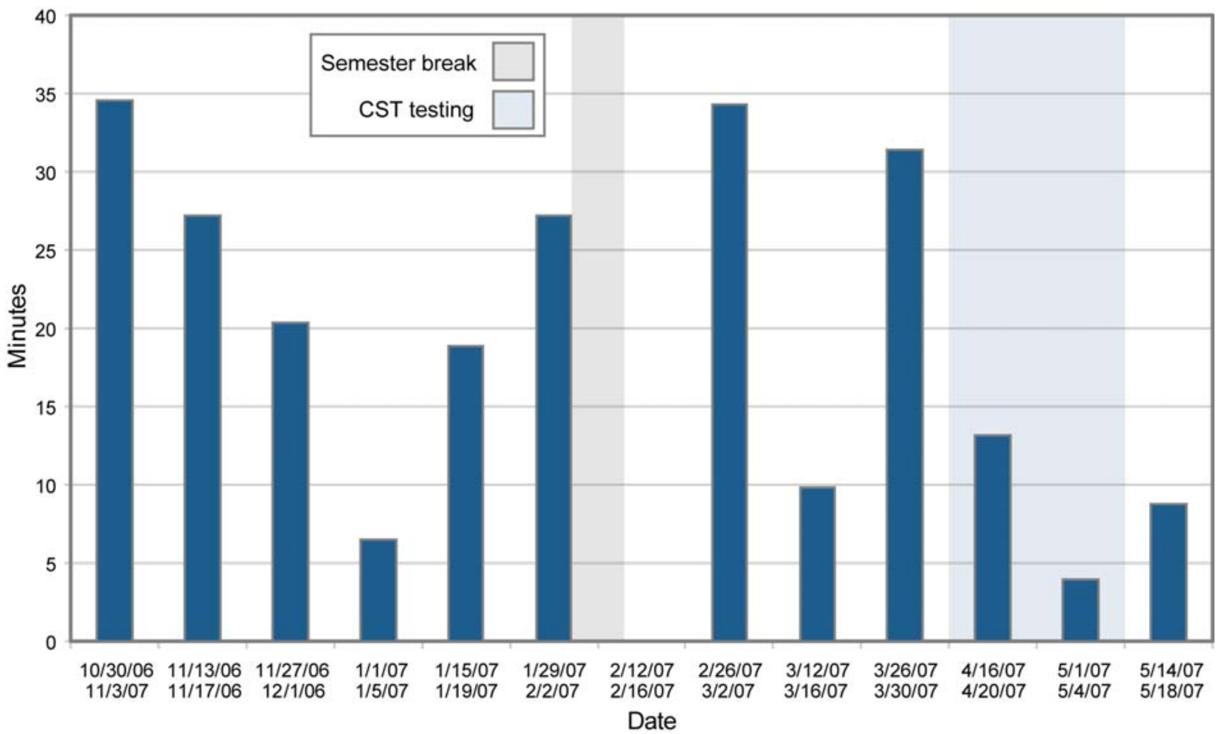


Figure 2. Average TI-Navigator Use in Minutes Per Week Per Class Period

On average, control classrooms used the graphing calculator for 26.6 minutes per week per class period. Teachers were surveyed every other week on the amount of time spent using the graphing calculators. Time spent using the graphing calculator indicates the amount of classroom exposure to the graphing calculator technology.

The following graph plots the average time spent using the graphing calculator based on survey data over the course of the study. The graph loosely follows a cyclical pattern with the highest use at the beginning of December and at the end of the year after state testing.

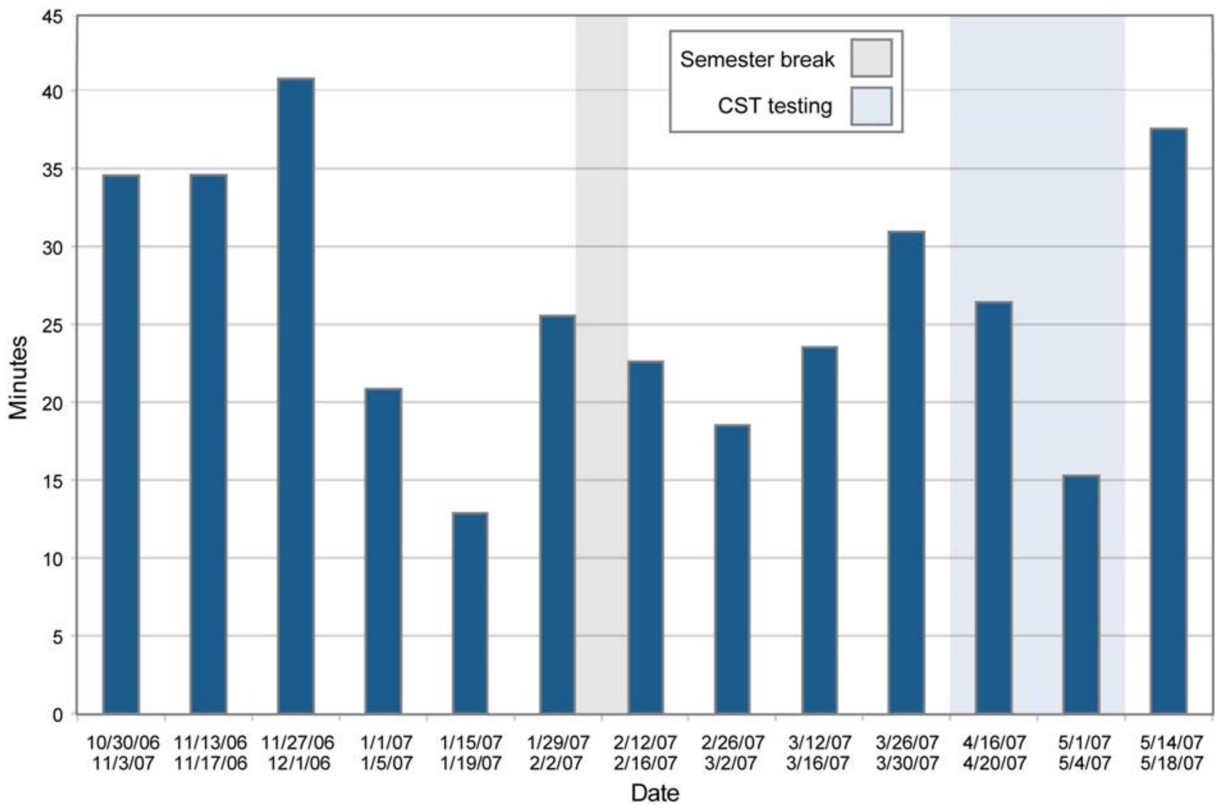


Figure 3. Average Graphing Calculator Use in Minutes Per Week Per Class Period

Use and Content Covered During Instruction

As shown in Figure 4, at each point surveyed, between two and five *GC+Nav* teachers perceived TI-Navigator as essential to their lessons. By comparison, Figure 5 shows that a majority of control teachers perceived the graphing calculators as essential to their lessons.

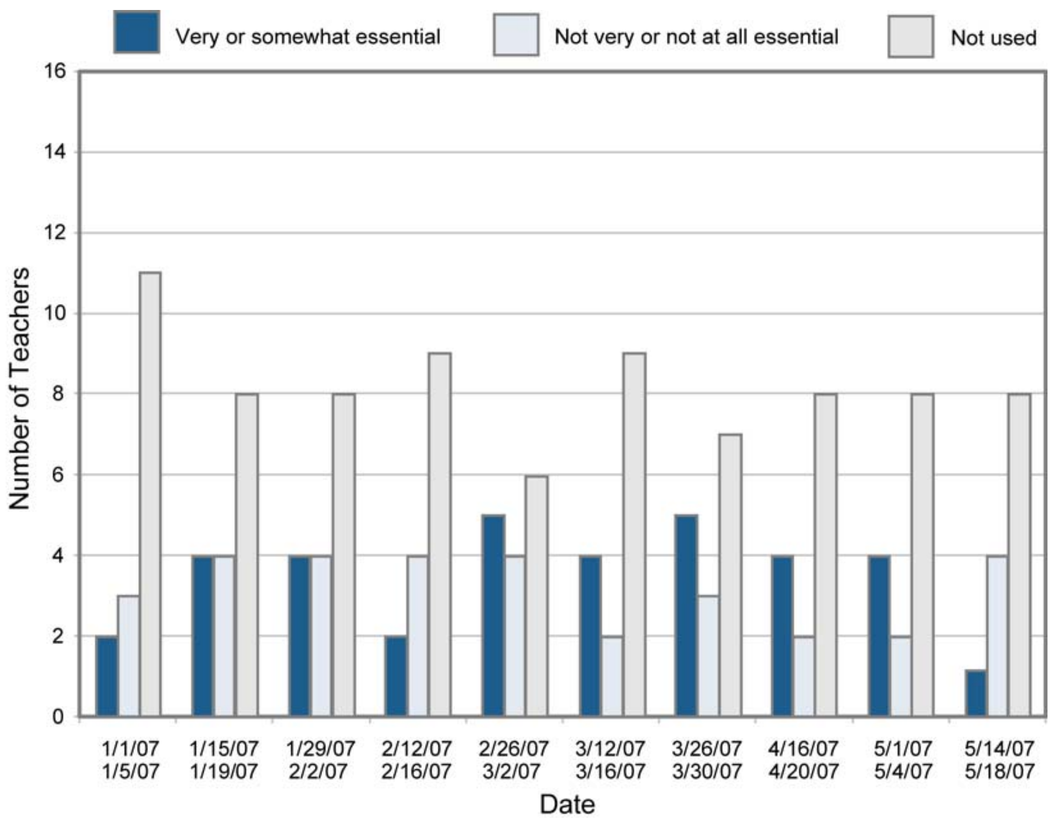


Figure 4. Do You Perceive TI-Navigator As Essential to Your Lesson?

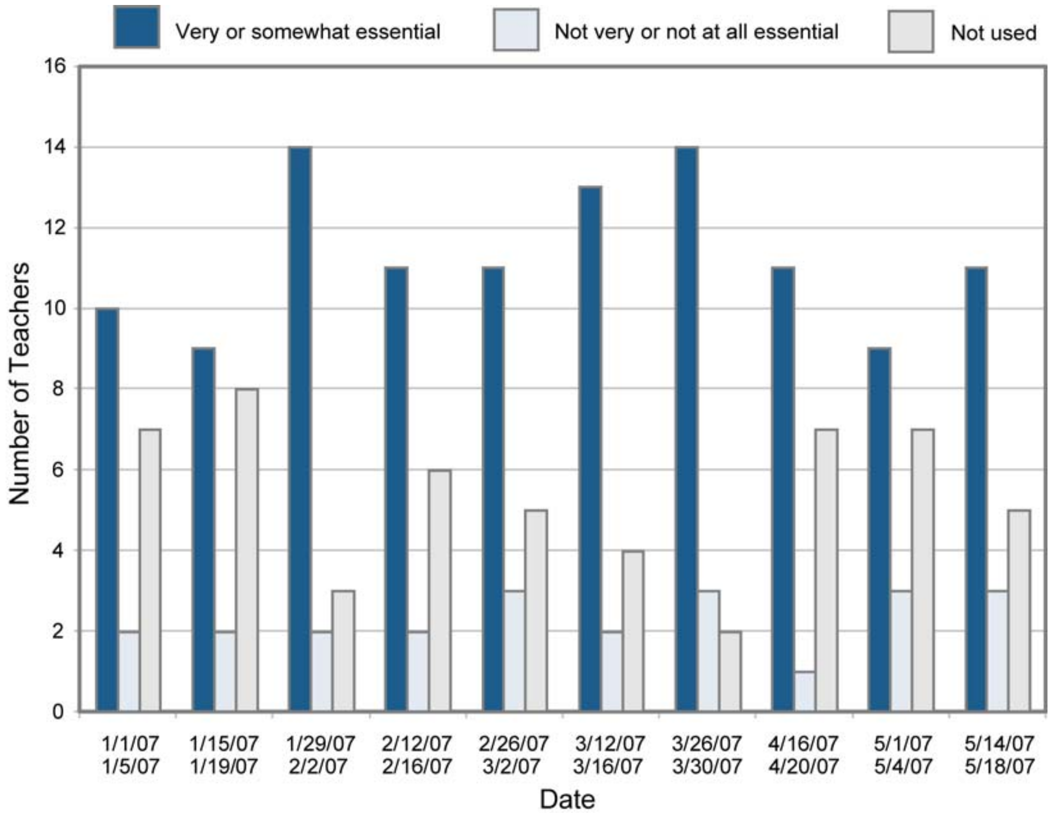


Figure 5. Do You Perceive Graphing Calculators As Essential to Your Lesson?

Table 23 shows that the most common use of TI-Navigator in December was for giving spontaneous LearningCheck quizzes. In February, the number of classrooms that reported using the system for this activity dropped from 11 to six. On both surveys, six classrooms reported using TI-Navigator to deliver classroom lessons.

Table 23. For What Tasks Do You Use the TI-Navigator System?

	Give formal tests/quizzes	Give spontaneous LearningCheck quizzes	Plan classroom lessons	Deliver classroom lessons	Distribute/ collect HW	Attendance/ class management
12/1/2006	1	11	3	6	1	2
2/2/2007	0	6	4	6	1	2
Note. 17 teachers responded.						

Table 24 shows the tasks for which the graphing calculator was used during instruction. The majority of teachers reported using the graphing calculator to plan and deliver classroom instruction. The control teachers who reported using the graphing calculators for attendance kept a classroom set of graphing calculators. The teachers could track attendance by which students had checked out calculators that class period.

Table 24. For What Tasks Do You Use the Graphing Calculator?

	Give formal tests/quizzes	Give spontaneous LearningCheck quizzes	Plan classroom lessons	Deliver classroom lessons	Distribute/ collect HW	Attendance/ class management
12/1/2006	9	7	13	13	1	2
2/2/2007	8	7	12	15	2	1

Note. 19 teachers responded.

Table 25 shows the number of *GC+Nav* teachers who used the graphing calculator alone or TI-Navigator to teach a specific algebra topic. Table 26 shows the number of control teachers who used the graphing calculator alone to teach a specific algebra topic. This number is displayed as a fraction of the total teachers who taught the topic. It appears that Algebra teachers chose to use the graphing calculator alone over using the TI-Navigator.

Table 25. Number of *GC+Nav* Teachers who used the Graphing Calculator only vs. TI-Navigator to Teach Specific Algebra Topics

Algebra topics	No. of teachers who used the graphing calculator alone	No. of teachers who used TI-Navigator
Real numbers	3/8	1/8
Solving linear equations	7/9	2/9
Graphing linear equations	8/9	1/9
Solving inequalities	3/9	1/9
Graphing inequalities	6/8	1/8
Exponents and exponential functions	5/8	2/8
Quadratic equations and functions	6/9	3/9
Polynomials and factoring	3/8	2/8
Rational equations and functions	2/5	1/5
Radicals	4/8	1/8
Probability and data analysis	0/0	0/0
Solving systems of equations	6/9	1/9
Data collection from the real world	0/0	0/0
Line of best fit for collected data	½	0/2
Analysis of collected data	0/0	0/0
Slope and intercept	7/9	4/9

Note. A total of 9 *GC+Nav* Algebra teachers responded.

Table 26. Number of Control Teachers who used the Graphing Calculator to Teach Specific Algebra Topics

Topics	Used graphing calculator to teach
Real numbers	8/13
Solving linear equations	9/13
Graphing linear equations	13/13
Solving inequalities	6/12
Graphing inequalities	9/12
Exponents and exponential functions	9/10
Quadratic equations and functions	10/11
Polynomials and factoring	4/11
Rational equations and functions	2/6
Radicals	7/9
Probability and data analysis	0/1
Solving systems of equations	10/10
Data collection from the real world	4/5
Line of best fit for collected data	7/8
Analysis of collected data	3/4
Slope and intercept	13/13

Note. A total of 13 control Algebra teachers responded.

Table 27 shows the number of *GC+Nav* teachers who used the graphing calculator alone or TI-Navigator to teach a specific geometry topic. Table 28 shows the number of controls teachers who used the graphing calculator alone to teach a specific geometry topic. This number is displayed as a fraction of the total teachers who taught the topic. It appears that teachers chose to use TI-Navigator in Geometry more than the teachers did in Algebra.

Table 27. Number of *GC+Nav* Teachers who used the Graphing Calculator only vs. TI-Navigator to Teach Specific Geometry Topics

Geometry topics	No. of teachers who used the graphing calculator alone	No. of teachers who used TI-Navigator
Segments, angles, bisectors, perpendiculars	6/6	3/6
Triangles-Relationships and congruence	4/6	3/6
Similarity-Triangles	3/5	3/5
Similarity-Polygons	2/5	2/5
Properties of quadrilaterals	3/5	2/5
Properties of polygons	2/4	1/4
Properties of circles	2/5	2/5
Transformations and tessellations	1/3	0/3
Measuring length and area	5/6	4/6
Surface area and volume-triangular solids	2/2	0/4
Surface area and volume-polygon solids	2/3	0/3
Surface area and volume-spheres	2/2	0/2
Trigonometry	4/5	2/5
Proofs	2/5	3/5

Note. A total of 7 *GC+Nav* Geometry teachers responded.

Table 28. Number of Control Teachers Who Used Graphing Calculators to Teach Specific Geometry Topics

Topics	Used graphing calculator to teach
Segments, angles, bisectors, perpendiculars	1/3
Triangles-Relationships and congruence	1/3
Similarity-Triangles	1/3
Similarity-Polygons	1/3
Properties of quadrilaterals	2/3
Properties of polygons	2/3
Properties of circles	2/3
Transformations and tessellations	1/2
Measuring length and area	2/3
Surface area and volume-triangular solids	0/1
Surface area and volume-polygon solids	0/1
Surface area and volume-spheres	0/1
Trigonometry	1/3
Proofs	0/3

Note. A total of 6 control Geometry teachers responded.

Use of Cabri Jr.

Cabri Jr. is a graphing calculator application designed exclusively for geometric constructions. Table 29 summarizes the geometry topics *GC+Nav* teachers reported using Cabri Jr. for when asked at two points. Table 30 summarizes the geometry topics control teachers reported using Cabri Jr. for when asked at two points.

Table 29. Number of *GC+Nav* Teachers Who Used Cabri Jr. to Teach Specific Geometry Topics

	1/5/2007	3/2/2007
Angle properties	3	2
Bisection		1
Circle properties	2	
Lines, parallel properties	3	2
Lines, perpendicular properties		1
Lines, segments, points		3
Measurement	2	1
Proofs		1
Quadrilateral properties	1	1
Transversals	1	
Triangle properties	4	2

Note. There were a total of 7 *GC+Nav* teachers who responded.

Table 30. Number of Control Teachers Who Used Cabri Jr. to Teach Specific Geometry Topics

	1/5/2007	3/2/2007
Angle properties	1	1
Bisection		
Circle properties		1
Lines, parallel properties	1	
Lines, perpendicular properties	1	
Lines, segments, points		
Measurement		
Proofs	1	
Quadrilateral properties	1	2
Transversals		
Triangle properties	1	3

Note. There were a total of 6 control teachers who responded.

GC+Nav and control geometry teachers reported using Cabri Jr. similarly. Although not used very regularly, Table 31 shows that all but one *GC+Nav* geometry teacher used the Cabri Jr. application. By comparison, Table 32 shows that control teachers appear to use the application more frequently than *GC+Nav* teachers.

Table 31. How Often Do You Use Cabri Jr. in One *GC+Nav* Geometry Class?

	4-5/week	2-3/week	Once a week	2-3/every other week	2-3/month	Never
1/5/2007	0	1	3	0	2	1
3/2/2007	0	0	2	1	3	1

Note. 7 *GC+Nav* teachers responded to both surveys.

Table 32. How Often Do You Use Cabri Jr. in One Control Geometry Class?

	4-5/week	2-3/week	Once a week	2-3/every other week	2-3/month	Never
1/5/2007	0	1	1	1	1	2
3/2/2007	0	1	0	1	2	2

Note. 6 control teachers responded to each survey.

When asked about projection devices and Cabri Jr. Table 33 shows that *GC+Nav* geometry teachers who used Cabri Jr. typically used the application with a projection device. All control geometry teachers who used Cabri Jr. used it with a projection device and about half almost always used a projection device as shown in Table 34. Control geometry teachers appear to use a projection device with Cabri Jr. more frequently between the two survey points suggesting that Cabri Jr. may have contributed to sharing work in geometry.

Table 33. How Often Do You Use a Projection Device with Cabri Jr. in One *GC+Nav* Geometry Class?

	Almost always (81-100%)	Frequently (61-80%)	Sometimes (41-60%)	Rarely (21-40%)	Almost never (0-20%)	I do not use Cabri Jr.
1/5/2007	4	1	0	0	1	1
3/2/2007	4	1	0	0	1	1

Note. 7 *GC+Nav* teachers responded to both surveys.

Table 34. How Often Do You Use a Projection Device with Cabri Jr. in One Control Geometry Class?

	Almost always (81-100%)	Frequently (61-80%)	Sometimes (41-60%)	Rarely (21-40%)	Almost never (0-20%)	I do not use Cabri Jr.
1/5/2007	2	0	0	2	0	2
3/2/2007	2	2	0	0	0	2

Note. 6 control teachers responded to each survey.

Use for Assessment

The following data indicate when and how TI-Navigator was used for assessment in the *GC+Nav* classrooms that implemented the system.

Three teachers reported that TI-Navigator was used for assessment purposes in January and four teachers in March. We present in Table 35 the responses from each of these teachers regarding the assessment frequency, the TI-Navigator function used, and the types of questions asked. Quick Poll with Yes/No or True/False questions were most commonly used. Two teachers used Open Ended questions. Of particular note is that only one teacher reports using an Activity Center style assessment during the study.

Table 35. Planned Formal Assessment

No. of teachers	Frequency	Quick poll	Learn check	Screen capture	Activity center	Yes-no / true-false	Multiple choice	Open ended
1/19/2007								
1	Weekly	•	•			•	•	•
1	Monthly	•	•			•	•	•
1	Monthly	•		•		•	•	
3/30/2007								
1	Weekly		•			•	•	•
1	Monthly	•	•			•	•	•
1	Monthly	•				•		
1	Monthly	•	•		•	•	•	•

Table 36. Casual Assessment

No. of teachers	Frequency	Quick poll	Learn check	Screen capture	Activity center	Yes-no / true-false	Multiple choice	Open ended
1/19/2007								
1	Weekly	•	•			•	•	•
1	Monthly	•	•			•	•	•
1	Monthly	•		•		•	•	
3/30/2007								
1	Weekly	•				•	•	•
1	Monthly	•	•			•	•	•
1	Monthly	•		•		•	•	
1	Monthly			•		•	•	•

Table 37 shows the average number of Quick Polls and LearningCheck administered at three times during the study. More *GC+Nav* teachers report using Quick Poll than LearningCheck when surveyed.

Table 37. During An Average Eligible Class Period, How Often Do You Administer Quick Polls and LearningCheck?

	Total no. of teachers	More than 5	4-5	2-3	1	Never	Other[a]
Quick Polls							
1/5/2007	17	4	1	3	0	8	1
3/2/2007	16	4	0	4	0	6	2
5/4/2007	15	1	2	4	3	4	1
LearningCheck							
1/5/2007	17	0	0	1	1	15	0
3/2/2007	16	0	0	0	3	10	3
5/4/2007	15	0	0	3	1	10	1

[a] Teachers provided no response or stated TI-Navigator was not used that week.

Teachers created their own LearningChecks rather than use the TI Activity Exchange for pre-made LearningChecks. Three teachers reported that LearningChecks found on the TI Activity Exchange were not entirely appropriate for their lessons and required comprehensive modification before they could be used in class. As a result, teachers believed that creating their own LearningChecks is more efficient.

The time required to create LearningCheck varied between 10 and 40 minutes. As shown in Table 37, classrooms that created a LearningCheck did not necessarily use the LearningCheck during class.

There are large differences in the number of LearningChecks that classrooms made:

- Three classrooms report making three
- One classroom reports making 20
- One classroom reports making more than 40 LearningChecks to coincide with each section of the Geometry textbook.

Graphing Calculators for Assessment

When first surveyed in January, two control teachers reported administering planned assessments that required the use of the graphing calculator beyond a simple calculation. These assessments were administered on a weekly or monthly basis and used only open-ended questions.

When surveyed again at the end of March, five control teachers reported administering planned assessments on a weekly or monthly basis using free-response questions. However nonsensical responses to this question by two teachers raise concerns about the question’s reliability. One teacher reported administering planned assessments, but selected “Never” as how often the assessments were administered. One classroom reported administering planned assessments on a monthly basis and selected “Other: After STAR” testing as the types of questions used.

When first surveyed in January, two control teachers report administering casual assessments that required the use of the graphing calculator beyond a simple calculation. These casual assessments were administered on a daily, weekly, and monthly basis using open-ended questions only.

When surveyed again at the end of March, one control teacher reported administering weekly casual assessments using open-ended questions only. One other teacher reported using the Activity Exchange website to create monthly casual assessments using open-ended questions only.

Cooperative Learning

The classroom technology deployed in the study has the potential to create a cooperative learning environment. A cooperative learning environment allows students to work in groups and pairs to share and comment on each other's work. The shared work can prompt discussions where students can lead instruction. The following data attempts to measure the degree of collaboration that occurs in the *GC+Nav* and control classrooms.

The use of projection devices with TI-Navigator or the graphing calculator, particularly TI-SmartView and Document Cameras, allow for the ability to project student work to the entire class and the potential for students comment on each other's work. Table 38 shows the types of projection devices used in *GC+Nav* classrooms and typically control classrooms. In both groups the most common projection devices used were overhead projectors.

Table 38. Types of Projectors Used in Classrooms

	Total no. of teachers	Overhead projector #	TI-SmartView #	TV #	Document camera #	I don't use a projector #	Other[a] #
***GC+Nav* classrooms**							
11/3/2006[b]	17	8	12	0	3	2	0
1/5/2007	17	6	4	1	1	3	2
3/2/2007	16	4	7	1	2	1	1
Control classrooms							
11/3/2006[b]	19	9	8	2	2	0	1
1/5/2007	19	7	6	2	1	2	1
3/2/2007	19	8	4	4	1	1	1

[a] Teachers in the other category did not provide responses to the questions.

[b] Teachers were allowed to select all devices that applied.

The classroom layout indicates the potential for collaboration between students. Students seated in small groups or pairs have an easier way to work with each other than if seated in individual rows. As shown in Table 39, more *GC+Nav* classrooms reported using pairs or small groups when teaching with TI-Navigator than without at both survey points. Control classrooms shown in Table 40 typically used rows of student desks.

Table 39. Classroom Layout

		Rows	Pairs/Small groups	Circle/Large group	Not used	Unspecified
		#	#	#	#	#
12/1/2006	**Teach with TI-Navigator**	4	12	0	1	0
	Teach without TI-Navigator	9	8	0	0	0
2/2/2006	**Teach with TI-Navigator**	5	10	0	2	0
	Teach without TI-Navigator	10	6	0	0	1

Note. 17 teachers responded to each survey.

Students in *GC+Nav* classes shown in Table 40 and Table 41primarily used TI-Navigator for paired and small group assignments and to demonstrate to the whole class at both times surveyed. Table 42 shows control classrooms used the graphing calculator primarily for individual tasks.

Table 40. When You Are Teaching, How Are the Seats In Your Classroom Organized?

	Rows	Small groups	Circle/ Large group	No response
12/1/2007	12	7	1	1
2/2/2007	11	6	1	1

Note. 19 teachers responded to each survey. Teachers could select more than one configuration.

Table 41. How Have Your Students Used TI-Navigator during class?

	Demonstrate to whole class	Individual assignments	Individual test/quiz	Paired/ small group assignments	Paired/ small group tests/ quizzes
12/1/2006	5	3	1	6	1
2/2/2007	6	1	1	3	1

Note. 17 teachers responded to each survey.

Table 42. How Have Your Students Used Graphing Calculators During Class?

	Demonstrate to whole class	Individual assignments	Individual test/quiz	Paired/ small group assignments	Paired/ small group tests/ quizzes
12/1/2006	6	16	12	8	3
2/2/2007	6	16	13	13	4
Note. 19 teachers responded to each survey.					

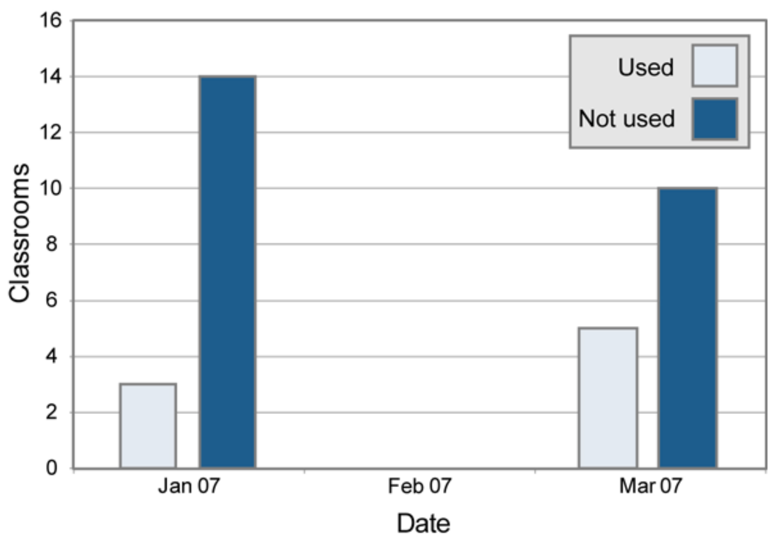


Figure 6. Classrooms Using Activity Center

Activity Center allowed students to plot points and lines on a shared coordinate plane. Figure 6 represents teacher survey responses. Although responses indicate that the Activity Center was not used much, use of the application appeared to increase over time.

The Activity Center is mainly useful in algebra classes. One algebra teacher initially reported not using the Activity Center but reported using it at least once a month when the question was asked again. Two algebra teachers reported using the Activity Center when it fit with lessons, such as graphing. Two geometry teachers that did not use the Activity Center stated the application was more appropriate for algebra.

Barriers to *GC+Nav* Implementation

During surveys, interviews, and classroom visits, teachers were asked to freely barriers to implementing the TI-Navigator. Teachers' free-responses were compiled and the common themes are summarized below. The following table presents teachers' comments regarding TI-Navigator hardware and software.

Table 43. Teachers' Comments Regarding TI-Navigator Hardware and Software

No. of teachers	Comments
Assembly and set up	
8	Need more time to practice using TI-Navigator
5	TI-Navigator needs to be free of glitches
3	More support and collaboration with other teachers using TI-Navigator
3	No cords running between desks and rows
1	More reliable technical support
1	TI-Navigator required more power outlets than the classroom can accommodate
2	Needs a permanent secure place for the equipment
TI-Navigator use	
4	Easier way to check that operating system and applications are up to date
4	Easier login process for students
3	Need to establishing a TI-Navigator routine in the classroom
3	Graphing calculators need fresh batters to work with TI-Navigator
1	Need a larger room with tables
2	Need block scheduling for more time to use TI-Navigator
Applications	
5	Premade and applicable LearningCheck that fit with the textbook
2	Typing Quick Poll questions out became wasted time. The same assessment purpose could be accomplished without TI-Navigator.
3	Students do not always take Quick Poll seriously and students would respond with disruptive answers.
1	Activity Center is too time consuming for the class.
3	Activity Center is more appropriate for algebra than geometry.
1	It is difficult to remember how to use Activity Center.
Projection devices	
3	TI-SmartView is too slow.
1	Keystroke history is distracting to class.
1	TI-SmartView needs to be magnified to be visible
Curriculum	
3	Conflict because calculators are banned from the California Standards Test.
2	Classrooms did not want to sacrifice instructional time to use TI-Navigator.

Summary of Implementation in *GC+Nav* and Control Groups

Every member of the *GC+Nav* group participated in at least part of the training offered by the study and five members of the control group did not participate in the graphing calculator training. There were differences in the number of teachers that actually used the technology in the classroom. The difference in usage may be attributable to each group's experience using the classroom technology. Half of the *GC+Nav* group had one semester of TI-Navigator experience while the other half was using TI-Navigator for the first time. By contrast, most of the control group had multiple years of experience teaching with graphing calculators. The control group could draw on this past experience to implement the graphing calculators.

Although teachers indicated they had the necessary resources to use the classroom technology at the beginning of each semester, teacher comments throughout the study reveal that crucial resources were missing from the *GC+Nav* group. The fact that half the *GC+Nav* teachers implemented TI-Navigator and only one *GC+Nav* teacher implemented TI-Navigator continuously raises the question of how TI-Navigator implementation could have been better supported. *GC+Nav* teachers most commonly cited technical problems and glitches in the system that prevented TI-Navigator use in the classroom. The initial problems so turned *GC+Nav* teachers off from using TI-Navigator that they were never able to continuously use the system and establish the classroom routine they recognized as being crucial for TI-Navigator use.

By comparison, every control classroom implemented the graphing calculator. The graphing calculator did not experience the technical problems as were reported with TI-Navigator. While control classrooms cite problems with the TI-SmartView projection, control classrooms had backups, such as document cameras and overhead projectors with TI-ViewScreens to substitute. Additional devices were not necessary for the graphing calculator to be used.

For TI-Navigator to be used successfully, the system must work out of the box as the graphing calculator does. Teachers and students cannot afford the time spent trouble shooting the system during classroom instruction as five teachers report glitches being a barrier to implementation and two teachers report problems at the beginning of the year prevented implementation from continuing.

The average time spent using the technology indicates the *GC+Nav* teachers exposed their students to TI-Navigator more than control teachers exposed their students to the graphing calculators. Even though only half of the *GC+Nav* classrooms used TI-Navigator, all but two *GC+Nav* classrooms report using the graphing calculators. While all control teachers used the graphing calculators, surveys indicated the *GC+Nav* group exposed their classrooms to the technology more than the control.

The *GC+Nav* group mainly used TI-Navigator to deliver classroom lessons and give spontaneous assessments, such as Quick Polls and LearningCheck. TI-Navigator was readily adopted for assessments given the Quick Poll, LearningCheck, and Screen Capture functions that *GC+Nav* teachers report using for assessments. *GC+Nav* teachers who used TI-Navigator were observed to have used Quick Poll on multiple occasions for multiple choice and open ended questions. However it is unclear to what extent teachers used the assessment data during instruction. It was observed that all *GC+Nav* teachers that used TI-Navigator with a projection device employed the assessments as modeled in the training: the teacher projected the various student solutions and the teacher made slight adjustments to re-teach or move on given the assessment results. However it was not observed or surveyed how the teacher used the classroom assessment data beyond that. The trainings noted that assessments results could be saved and reviewed with individual students after class. However to use TI-Navigator in this capacity, teachers need to be trained on how to efficiently interpret and use the assessment data to target specific student needs.

By comparison, the graphing calculator did not have built in functions for assessment similar to TI-Navigator. Few control classrooms reported administering assessments that required the graphing calculator beyond a simple computation. Graphing calculators were most commonly used to plan and deliver classroom instruction.

Creation of cooperative learning environments is important because it gives students leadership over their learning. When surveyed, most teachers in the *GC+Nav* group and control groups used a projection device while teaching. The use of projection devices in both groups indicates that the potential for students to engage in cooperative work exists. With projection devices available, student work can be easily shared with the class that could allow students to comment and prompt classroom discussion. It is unclear whether teachers organized the learning environment to make full use of the potential. Physical classroom arrangement is yet another indicator of a cooperative learning environment.

Students arranged in small groups or pairs appear more likely to collaborate than in rows of individual desks. When using TI-Navigator, *GC+Nav* classrooms report using a small group or paired layout, while most control classrooms report teaching to students seated in rows.

The student use of the technology also indicates the amount of student involvement in learning. In *GC+Nav* classrooms, students used TI-Navigator for collaborative activities as compared to students in control classrooms who used the graphing calculators for individual activities. *GC+Nav* teachers report students using TI-Navigator to demonstrate to the entire class or when performing paired or small group activities. Control classrooms reported students using the graphing calculators for individual assignments or individual tests and quizzes.

GC+Nav Implementation Rating

Using the data collected from TI-Navigator log files, surveys, and classroom observations, we rated classrooms on TI-Navigator implementation using a three-point scale. The scale was divided as *No-Implementation*, *Limited-Implementation*, and *Comprehensive-Implementation*.

Classrooms that did not use TI-Navigator beyond the trainings were rated as *No-Implementation*. Classrooms that demonstrated TI-Navigator use on more than one occasion were classified as *Limited-Implementation*. Of the classrooms that used TI-Navigator, classrooms were designated as *Comprehensive-Implementation* based on the frequency of TI-Navigator use, and the extent to which it was used for formative assessment and sharing class work.

Table 44 shows the number of classroom that fell in each of the categories for Algebra and Geometry.

Table 44. Number of Teachers in Each Implementation Rating

	No-Implementation	Limited - Implementation	Comprehensive-Implementation
Algebra			
GC+Nav A	2	0	0
GC+Nav B	2	4	0
Geometry			
GC+Nav A	1	2	3
GC+Nav B	1	0	0

Student-level Impact Results

Overview

The primary goal of our experiment was to understand the impact of TI-Navigator on student algebra and geometry achievement.

In the following sections, our examination of the impact follows a similar pattern.

1. Program impact on students: We will first address the impact on algebra achievement and then, the impact on geometry achievement. Within each content area, we show whether there is a difference in average performance on the posttest between students in *GC+Nav* and control classes. We also show whether there is a difference in performance between the two conditions for a student with an average score on the pretest.
2. Moderation of the impact: For each outcome scale, we then examine whether the impact of the program is different depending on a moderator, which is a condition or characteristic that existed prior to the program at the student or teacher level. We always begin by examining whether the impact of the program differs depending on the students' pretest scores—do pretest scores moderate the impact?

Impact of TI-Navigator on NWEA Algebra

Overall Score on the NWEA Algebra Test

We first address algebra outcomes using the NWEA Algebra scale. Table 45 provides a summary of the sample we used and the results for the comparison of NWEA Algebra scores for students in *GC+Nav* and control groups. The "Unadjusted" row gives information about all the students in the original sample for whom we have a pretest and posttest. This shows the means and standard deviations as well as counts for students, classes, and teachers in that group. The last two columns provide the effect size, that is, the size of the difference between the means for *GC+Nav* and control in standard deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference as large as, or larger than, the absolute value of the one observed when there truly is no difference. The "Adjusted" row is based on the same sample of students. The means, and therefore the effect size, are adjusted to take into account the student pretest scores; hence, these statistics are adjusted for any chance imbalances on the pretest between the two randomized groups[6].

[6] All statistical equations that follow build on the one used to compute the adjusted effect size.

Table 45. Overview of Sample and Impact of TI-Navigator on NWEA Algebra

	Condition	Means	Standard deviations[a]	No. of students	No. of classes	No. of teachers	Effect size	*p* value[b]	Percentile standing
Un-adjusted effect size	**Control**	-0.01	1.04	424	29	13	0.04	0.97	1.60%
	GC+Nav	0.03	0.95	229	20	8			
Adjusted effect size	**Control**	-0.01	1.04	The same sample is used in both calculations			-0.15	0.43	-5.96%
	GC+Nav	-0.16[c]	0.95						

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the scores of the students in the sample for that row.

[b] The *p* value for the unadjusted effect size is computed using a model that figures in clustering of students in teacher level but does not adjust for any other covariates. The *p* value for the adjusted effect size is computed using a model that controls for clustering and that includes both the pretest and, where relevant, indicators for upper-level units (pairs) within which the units of randomization are nested.

[c] Modeling separate intercepts for upper-level units leads to estimates of performance, in the absence of treatment, which are specific to those units. For purposes of display, to set the performance estimate for the control group, we compute the average performance for the sample of control cases used to calculate the adjusted effect size. The estimated treatment effect, which is constrained to be constant across upper-level units, is added to this estimate to show the relative advantage or disadvantage to being in the treatment group.

Figure 7 provides a visual representation of results from the row labeled "Adjusted" in Table 45. The bar graphs represent average performance using the metric of NWEA Algebra. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students' pretest scores and other fixed effects. The overall algebra effect size (in standard deviation units) is -.15, which is equivalent to a loss of 6.0 percentile points for the median control group student if the student had received TI-Navigator. The high *p* value for the treatment effect (.43) indicates we should have no confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The overlap in these intervals further indicates that any difference we see is easily due to chance.

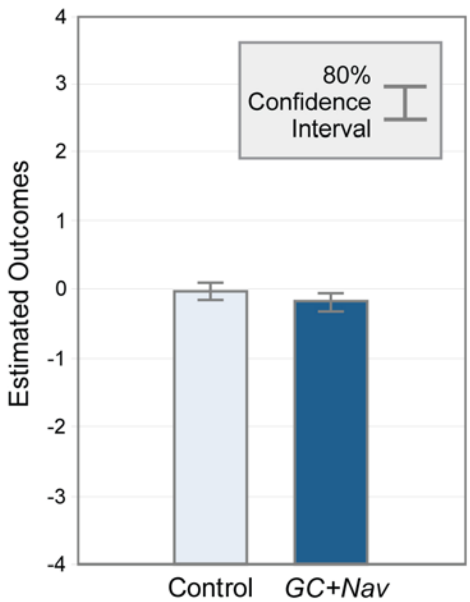


Figure 7. Impact on NWEA Algebra Achievement: Adjusted Means for Control and *GC+Nav*

Moderating Variables[7]

We now report on our examination of the moderating effects of other variables (performance on pretest, gender, and English proficiency). We provide a separate table of results for each of these moderator analyses. The fixed factor part of each table provides estimates of the effects of interest.

Including Pretest as a Moderator

We first show whether the impact of TI-Navigator is different for students at different levels of prior achievement. At the bottom of the table we give results for technical review—these often consist of what are called random effects estimates. As was described earlier in this report, random effects are added to the statistical equation to account for dependencies in observed scores that happen because students come from the same classes or teachers[8].

---

[7] As a rule, we decide on moderator variables before the beginning of the experiment. We declare which variables we will examine the moderating effects of in advance to demonstrate that we are not mining results post hoc. The moderators are variables of theoretical interest that potentially affect how strongly the treatment impacts the outcome. We graph only the result for which we have at least limited confidence that the true effect is different from zero.

Statistical power for detecting moderating effects of variables is lower than for detecting the average effect because the former essentially looks at more subtle differences in outcomes (differences, between subgroup, in the difference in posttest performance between treatment and control). The cost to power is greater if the moderator is at the level of randomization than if it is at levels below the level of randomization.

[8] In some cases, to account for these dependencies, we model fixed rather than random effects but do not present the individual fixed effects estimates in the table.

Table 46 shows the estimated impact of TI-Navigator on the performance of students with an average pretest as measured by NWEA Algebra as well as the moderating effect of their prior scores.

Table 46. Impact of TI-Navigator on NWEA Algebra Achievement

Fixed effects[a]	Estimate	Standard error	DF	t value	*p* value
Outcome for the control student with an average pretest	0.05	0.35	6	0.14	.89
Change in outcome for the control student for each unit-increase on the pretest	0.76	0.04	630	20.18	<.01
Effect of *GC+Nav* for a student with an average pretest	-0.15	0.17	6	-0.91	.4
Change in the effect of *GC+Nav* for each unit-increase on the pretest	0.08	0.07	630	1.14	.25
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.08	0.06		1.43	.08
Within-teacher variation	0.4	0.02		17.75	<.01

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

The row in the table labeled "Effect of *GC+Nav* for a student with an average pretest" tells us whether TI-Navigator made a difference in the NWEA Algebra scores for a student who has an average score on the pretest. The estimate associated with *GC+Nav* is -0.15. This shows a small negative difference associated with *GC+Nav*. The *p* value of.4 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 40% of the time when there truly is no effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero.

We also estimated the moderating effect of the pretest score on the impact of *GC+Nav* (row 4) to determine whether the intervention was differentially effective for students at different points along the pretest scale. The coefficient associated with the interaction of pretest with treatment is .08, which shows a small increase in the treatment effect with each one-unit increase on the pretest. The *p* value of .25 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 25% of the time when in fact there is zero impact[9]. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true differential impact is different from zero. In other words, the effect of TI-Navigator was the same for students, regardless of how well a student performed on the pretest.

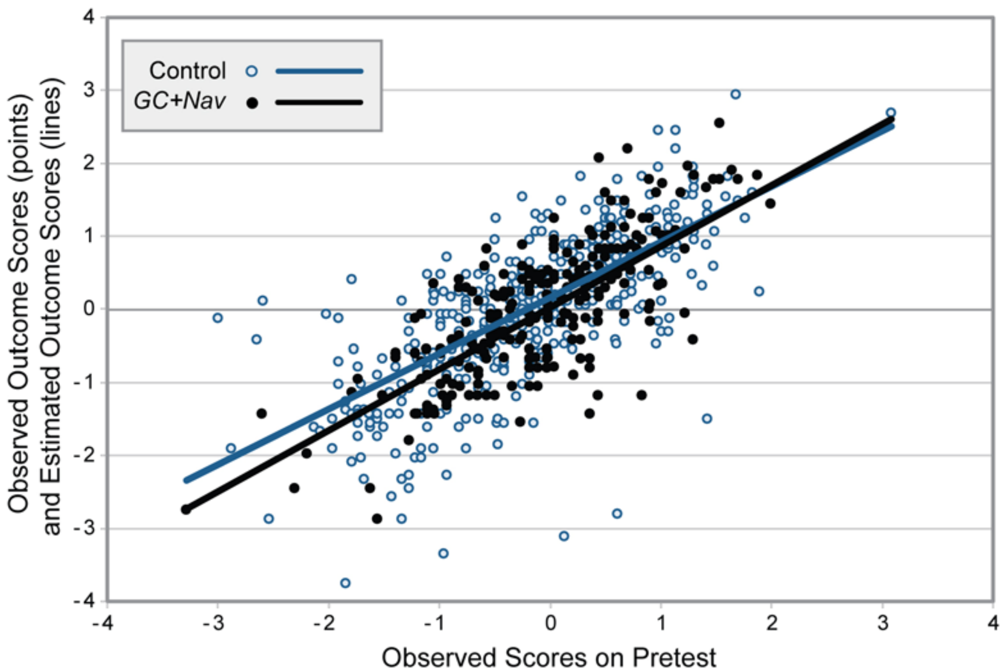


Figure 8. Comparison of Estimated and Actual Outcomes for *GC+Nav* and Control Group Students (NWEA Algebra Achievement)

As a visual representation of the result described in Table 46, we present a scatterplot in Figure 8, which shows student performance at the end of the year in Algebra as measured by the NWEA Algebra test, against their performance on NWEA in the fall. This graph shows where each student started in terms of his or her pretest score (horizontal x-axis) and his or her outcome score[10]. We remind the reader that pretest scores have been z-transformed within test-type. Each point plots one student's post-intervention score against his or her pre-intervention score. The darker points represent *GC+Nav* students; the lighter points, control students.

[9] In the model used, intercepts are modeled as random at the teacher level and as fixed at the pair level; however, slopes are not modeled as random; the interaction of pretest with treatment and the corresponding *p* value do not reflect uncertainty due to a potential re-sampling of teachers.

[10] The value along the horizontal scale for each point represents the distance from the origin, in standard deviation units.

The two lines are the estimated values on the posttest for students in the *GC+Nav* and control conditions. We see very little difference in impact across the prior score scale.[11]

Including Gender as a Moderator

In addition to looking at the main effect of TI-Navigator, and the differential effect of pretest, we estimated the interactions of *GC+Nav* and gender of the students. In particular, we were interested in whether the condition's effect was different for girls and boys. Table 47 shows the moderating effect of gender on students' performance on NWEA Algebra. The advantage of being in the *GC+Nav* condition is slightly more for girls than it is for boys. The *p* value of .33 means we have no confidence that the actual differential impact is different from zero.

Table 47. Moderating Effect of Gender on NWEA Algebra

Fixed effects [a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the control girl with an average pretest	-0.02	0.36	6	-0.06	.95
Change in outcome for each unit-increase on the pretest	0.78	0.03	629	24.87	<.01
Control group difference (boys minus girls) in the outcome	0.08	0.06	629	1.24	.22
Effect of *GC+Nav* for girls	-0.1	0.18	6	-0.53	.62
Net difference (boys minus girls) in the effect of *GC+Nav*	-0.1	0.11	629	-0.97	.33
Random effects [b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.09	0.06		1.46	.07
Within-teacher variation	0.4	0.02		17.74	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. 653 students were used in this moderator model, there are no influential points.

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

---

[11] When we limit the analysis to only those students who took MAP posttest, the *p* value for the interaction drops to .01, with *GC+Nav* students outperforming controls at the high end of the pretest scale. Pairs were modeled as a fixed factor, resulting in a separate intercept estimate for each pair. To fix the vertical location of the prediction lines, we selected the median estimate from among the pair estimates for the intercept.

Including English Proficiency as a Moderator

We were also interested in the moderating effect of student English proficiency. In particular, we wanted to know whether TI-Navigator was differentially effective for English proficient students and for English learners. Table 48 shows the results for the moderating effect of student English proficiency. The advantage of being in the *GC+Nav* condition is slightly more for English proficient students. The *p* value of .60 means we have no confidence that the actual differential impact is different from zero[12].

Table 48. Moderating Effect of English Proficiency on NWEA Algebra

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the Non-English proficient control with an average pretest	0.07	0.37	6	0.18	.86
Change in outcome for each unit-increase on the pretest	0.8	0.03	618	24.19	<.01
Control group difference (English proficient minus not proficient) in the outcome[c]	-0.06	0.08	618	-0.83	.41
Effect of *GC+Nav* for Non-English proficient student	-0.19	0.2	6	-0.98	.37
Net difference (English proficient minus not proficient) in the effect of *GC+Nav*	0.07	0.13	618	0.53	.6
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.09	0.06		1.47	.07
Within-teacher variation	0.39	0.02		17.58	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 643 students we have information on this moderator, we removed 1 because they were influential points or outliers. Hence, 642 students were used in the impact model.

Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers..

The estimate of no difference between English-speakers and English-learners among controls is due to the inclusion of pretest in the model, which makes the 'English-speaker effect be 'net of' pretest. When we exclude pretest from the model, the average for English-speakers is larger than the average for non-English speakers.

[12] When we limit the analysis to only those students who took MAP posttest, the *p* value for the interaction drops to .16, with a slightly higher treatment effect for English proficient students.

Impact of TI-Navigator on CST Algebra

Overall Score on the CST Algebra Test

We now address algebra outcomes using the CST Algebra scale. Table 49 provides a summary of the sample we used and the results for the comparison of CST Algebra scores for students in *GC+Nav* and control groups. The "Unadjusted" row gives information about all the students in the original sample for whom we have a pretest and posttest. This shows the means and standard deviations as well as counts for students, classes, and teachers in that group. The last two columns provide the effect size, that is, the size of the difference between the means for *GC+Nav* and control in standard deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference as large as, or larger than, the absolute value of the one observed when there truly is no difference. The "Adjusted" row is based on the same sample of students. The means, and therefore the effect size, are adjusted to take into account the student pretest scores; hence, these statistics are adjusted for any chance imbalances on the pretest between the two randomized groups.[13]

Table 49. Overview of Sample and Impact of TI-Navigator on CST Algebra

	Condition	Means	Standard deviations[a]	No. of students	No. of classes	No. of teachers	Effect size	*p* value[b]	Percentile standing
Un-adjusted effect size	**Control**	294.26	43.11	453	28	13	-0.09	0.93	-3.59%
	GC+Nav	290.23	45.90	279	20	8			
Adjusted effect size	**Control**	294.26	43.11	The same sample is used in both calculations			-.26	0.35	-8.32%
	GC+Nav	282.81[c]	45.90						

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the scores of the students in the sample for that row.

[b] The *p* value for the unadjusted effect size is computed using a model that figures in clustering of students in teacher level but does not adjust for any other covariates. The *p* value for the adjusted effect size is computed using a model that controls for clustering and that includes both the pretest and indicators for upper-level units within which the units (pairs) of randomization are nested.

[c] Modeling separate intercepts for upper-level units leads to estimates of performance, in the absence of treatment, which are specific to those units. For purposes of display, to set the performance estimate for the control group, we compute the average performance for the sample of control cases used to calculate the adjusted effect size. The estimated treatment effect, which is constrained to be constant across upper-level units, is added to this estimate to show the relative advantage or disadvantage to being in the treatment group.

[13] The goal of modeling the pretest and pair effects is to increase the precision of the estimate of the program effect by accounting for variation due to these factors. These effect estimates are usually carried forward to all subsequent statistical equations that we use.

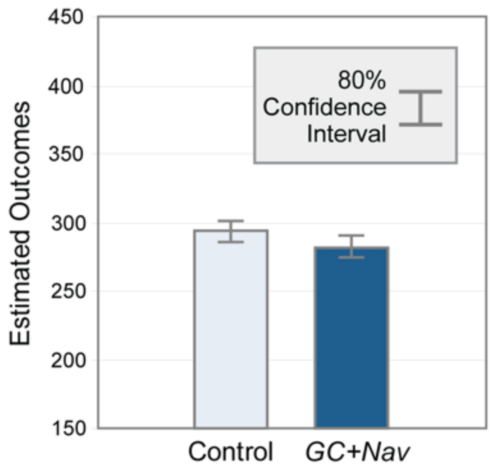


Figure 9. Impact on CST Algebra Achievement: Adjusted Means for Control and *GC+Nav*

Figure 9 provides a visual representation of results from the row labeled “Adjusted” in Table 49. The bar graphs represent average performance using the metric of CST Algebra. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students’ pretest scores and other fixed effects. The overall algebra effect size (in standard deviation units) is -.26, which is equivalent to a loss of 8.3 percentile points for the median control group student if the student had received TI-Navigator*.* The high *p* value for the treatment effect (.35) indicates we should have no confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The overlap in these intervals further indicates that any difference we see is easily due to chance.

Moderating Variables

We now report on our examination of the moderating effects of other variables (e.g., performance on pretest, gender, and English proficiency). We provide a separate table of results for each of these moderator analyses. The fixed factor part of each table provides estimates of the effects of interest.

Including Pretest as a Moderator

We first show whether the impact of TI-Navigator is different for students at different levels of prior achievement. At the bottom of the table we give results for technical review—these often consist of what are called random effects estimates. As was described earlier in this report, random effects are added to the statistical equation to account for dependencies in observed scores that happen because students come from the same classes or teachers[14].

Table 50 shows the estimated impact of TI-Navigator on the performance of students with an average pretest score as measured by CST Algebra as well as the moderating effect of their prior scores.

[14] In some cases, to account for these dependencies, we model fixed rather than random effects but do not present the individual fixed effects estimates in the table.

Table 50. Impact of TI-Navigator on CST Algebra

Fixed effects[a]	Estimate	Standard error	DF	t value	*p* value
Outcome for the control student with an average pretest	294.7	22.79	6	12.93	<.01
Change in outcome for the control student for each unit-increase on the pretest	24.98	1.84	708	13.56	<.01
Effect of *GC+Nav* for a student with an average pretest	-11.83	10.94	6	-1.08	.32
Change in the effect of *GC+Nav* for each unit-increase on the pretest	4.04	3.08	708	1.31	.19
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	366.7	233.92		1.57	.06
Within-teacher variation	988.65	52.53		18.82	<.01

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

The row in the table labeled "Effect of *GC+Nav* for a student with an average pretest" tells us whether TI-Navigator made a difference in the CST Algebra scores for a student who has an average score on the pretest. The estimate associated with *GC+Nav* is -11.83. This shows a small negative effect associated with *GC+Nav*. The *p* value of 0.32 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 32% of the time when there truly is no effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero.

We also estimated the moderating effect of the pretest score on the impact of *GC+Nav* (row 4) to determine whether the intervention was differentially effective for students at different points along the pretest scale. The coefficient associated with the interaction of pretest with treatment is 4.04, which shows a small) increase in the treatment effect with each one-unit increase on the pretest. The *p* value of .19 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 19% of the time when in fact there is zero impact[15].

[15] In the model used, intercepts are modeled as random at the student and teacher levels and as fixed at the pair level; however, slopes are not modeled as random; the interaction of pretest with treatment and the corresponding *p* value do not reflect uncertainty due to a potential re-sampling of teachers.

Using the criteria outlined earlier in the report, we conclude that we have limited confidence that the true impact is different from zero.

As a visual representation of the result described in Table 50, we present a scatterplot in Figure 10, which shows student performance at the end of the year in Algebra as measured by the CST Algebra test, against their performance on NWEA in the fall. This graph shows where each student started in terms of his or her pretest score (horizontal x-axis) and his or her outcome score[16]. We remind the reader that pretest scores have been z-transformed within test-type. Each point plots one student's post-intervention score against his or her pre-intervention score. The darker points represent *GC+Nav* students; the lighter points, control students.

The two lines are the estimated values on the posttest for students in the *GC+Nav* and control conditions. We see a slight difference in impact across the prior score scale. Consistent with the results described above, we observe that TI-Navigator and the programs used in the control classrooms were nearly equally effective as measured by CST Algebra.[17] [18]

---

[16] The value along the horizontal scale for each point represents the distance from the origin, in standard deviation units.

[17] Pairs were modeled as a fixed factor, resulting in a separate intercept estimate for each pair. To fix the vertical location of the prediction lines, we selected the median estimate from among the pair estimates for the intercept.

[18] We observe two features of the scatterplot. First, there appears to be heteroscedasticity - variance in the posttest is the not the same across the scale of the pretest. To address this we use the 'unstructured' option in SAS which leaves the fitted covariance matrix for observations unconstrained. Second, we observe that there is a slight curvilinear trend in the scatter – there seems to be a non-linear rise in the trend of the scatter as you move to the right. This is probably due to characteristics of the CST scale. Tests of skew and kurtosis in the outcome indicated that, according to conventional standards, there was no need to transform posttest scores. Nonetheless, we were concerned that that the estimate of the interaction effect may be driven by the fact that we fit straight-line effects on a curvilinear scatter. To check this possibility we also used a model that included a quadratic term for the pretest as well as a term for the interaction between the pretest-squared and treatment. When we modeled this curvature, the quadratic term was highly statistically significant and the p-value for the interaction rose to .25. This result and the fact that the p value for the interaction for the model without the quadratic term is .19 leads us to conclude that we have no confidence that there is a differential impact.

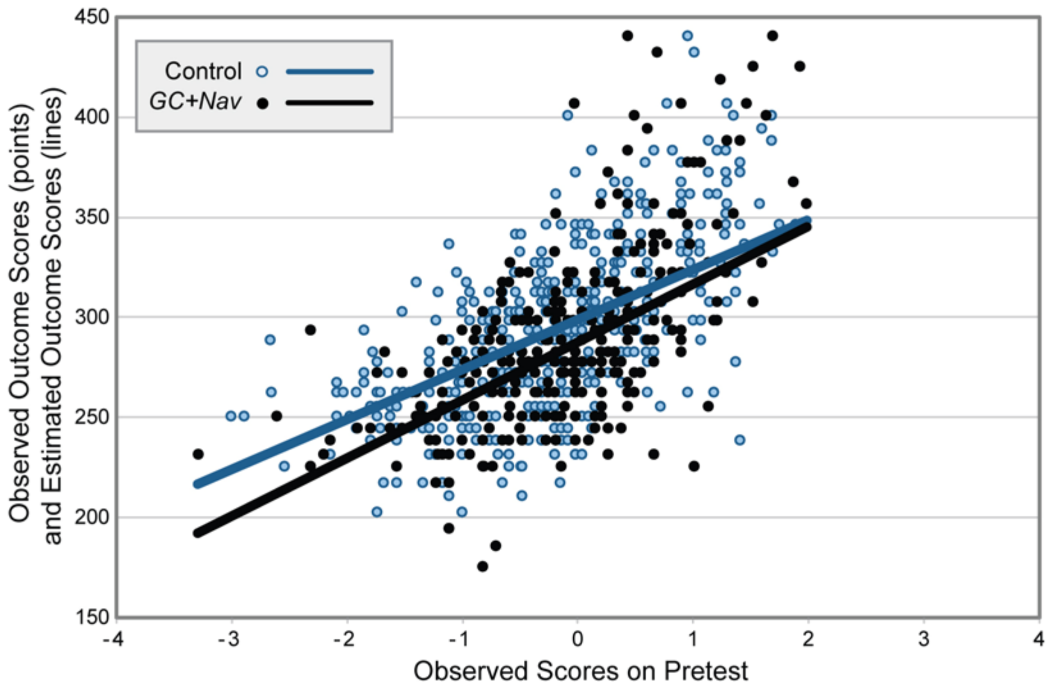


Figure 10. Comparison of Estimated and Actual Outcomes for *GC+Nav* and Control Group Students (CST Algebra Achievement)

Figure 11 shows the estimated difference between the *GC+Nav* and control groups for different points along the prior score scale. In this graph the estimated difference between *GC+Nav* and control groups is expressed as the straight line in the middle of the shaded bands—it is the estimated outcome for a *GC+Nav* student minus the estimated outcome for a control student. Around the difference line, we provide gradated bands representing confidence intervals. These confidence intervals are an alternative way of expressing uncertainty in the result. The band with the darkest shading surrounding the dark line is the "50-50" area, where the difference is considered equally likely to lie within the band as not. The region within the outermost shaded boundary is the 95% confidence interval—we are 95% sure that the true difference lies within these extremes. Between the 50% and 95% confidence intervals we also show the 80% and 90% confidence intervals. Consistent with the results in Figure 10, we see a weak trend of a differential impact of the intervention across the prior score scale as measured by CST Algebra, but given the high *p* value, we have limited confidence that prior score moderates the treatment effect. Considering the points representing the median student in the bottom and top quartiles, it appears that TI-Navigator has less benefit for the lower scoring students. However, neither point is sufficiently far from zero to give us confidence that it warrants a firm conclusion[19].

[19] The fact that the *p* value for the interaction rises above .20 when we model a quadratic term, is further indication that we should not conclude that there is a differential effect of the intervention. In the results of moderator analyses the standard error for the interaction shown in the table is generally not the same as the standard error used to express confidence intervals in the graphs. The standard error for the interaction expresses uncertainty in the parameter that measures a difference in impact (i.e., the difference in the difference

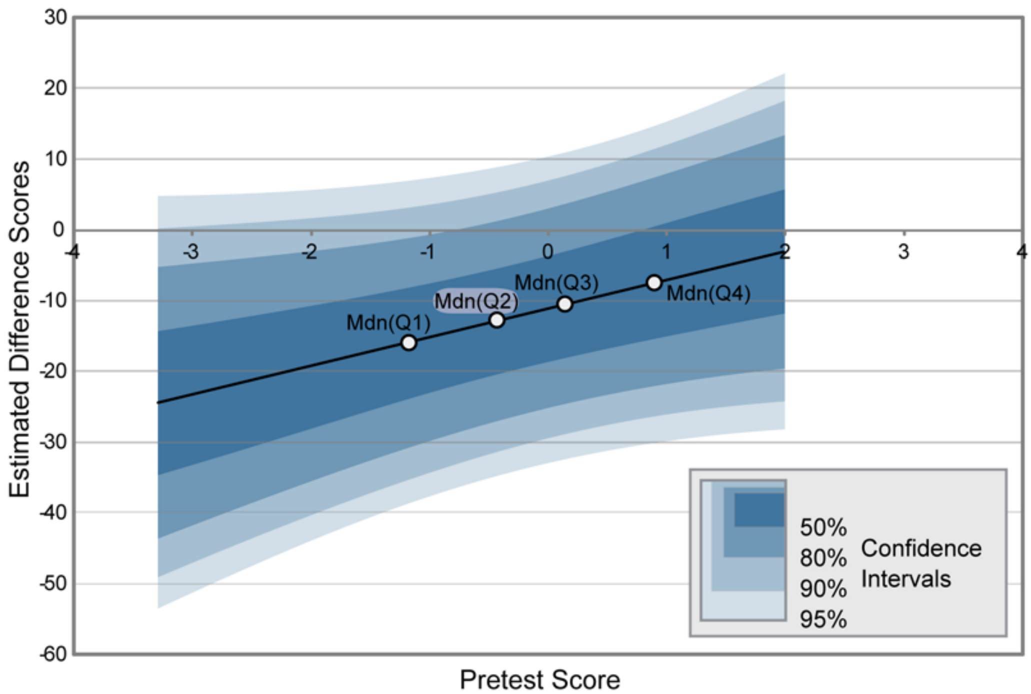


Figure 11. Differences between *GC+Nav* and Control Group CST Algebra Outcomes: Median Pretest Scores for Four Quartiles Shown

---

between treatment and control). The standard errors used for the confidence intervals in the graphs express uncertainty in the impact (i.e., the difference between treatment and control) at given levels of the moderator.

Including Gender as a Moderator

In addition to looking at the main effect of TI-Navigator, and the differential effect of pretest, we estimated the interaction of *GC+Nav* with gender. That is, we were interested in whether the condition's effect was different for girls and boys. Table 51 shows the moderating effect of gender on students' performance on CST Algebra.

Table 51. Moderating Effect of Gender on CST Algebra

Fixed effects	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a control girl with an average pretest	293.55	23.42	6	12.54	<.01
Change in outcome for each unit-increase on the pretest	26.45	1.49	707	17.74	<.01
Control group difference (boys minus girls) in the outcome	-1.56	3.01	707	-0.52	.6
Effect of *GC+Nav* for girls	-11.87	11.48	6	-1.03	.34
Net difference (boys minus girls) in the effect of *GC+Nav*	0.87	4.9	707	0.18	.86
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	388.98	246.04		1.58	.06
Within-teacher variation	991.6	52.73		18.81	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 732 students we have information on this moderator, we removed 1 because it was an influential point. Hence, 731 students were used in the moderator model.

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers..

We see that there is essentially no difference in the treatment effect between boys and girls. The *p* value for the moderating effect is .86, which gives us no confidence that the true differential effect is different from zero.

Including English Proficiency as a Moderator

We were also interested in the moderating effect of student English proficiency. In particular, we wanted to know whether TI-Navigator was differentially effective for English proficient students and for English learners. Table 52 shows the results for the moderating effect of student English proficiency.

We see a negative effect of the intervention for English speakers compared to non-English speakers. The *p* value for this result is .08; however, when we model a quadratic term to account for curvature in the scatter, the coefficient for the quadratic is highly significant and the *p* value for the interaction rises to .19, giving us limited confidence that the true moderating effect is different from zero. The moderating effect continues to be negative.

Table 52. Moderating Effect of English Proficiency on CST Algebra

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the Non-English proficient control with an average pretest	293.93	24.68	6	11.91	<.01
Change in outcome for each unit-increase on the pretest	27.21	1.53	694	17.78	<.01
Control group difference (English proficient minus not proficient) in the outcome[c]	-3.5	3.79	694	-0.92	.36
Effect of *GC+Nav* for Non-English proficient student	-3.94	12.48	6	-0.32	.76
Net difference (English proficient minus not proficient) in the effect of *GC+Nav*	-10.13	5.87	694	-1.73	.08
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	432.26	271.57		1.59	.06
Within-teacher variation	972.11	52.17		18.63	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 720 students we have information on this moderator; we removed 2 because they were influential points or outliers. Hence, 718 students were used in the moderator model.

Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

The estimate of no difference between English-speakers and English-learners among controls is due to the inclusion of pretest in the model, which makes the 'English-speaker effect' be 'net of' pretest. When we exclude pretest from the model, the average for English-speakers is larger than the average for non-English speakers.

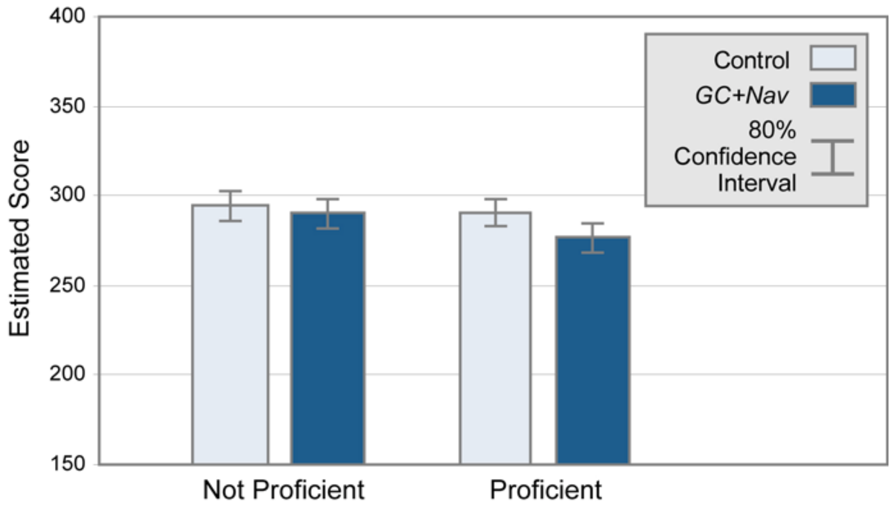


Figure 12. Moderating Effect of English Proficiency on CST Algebra

Impact of TI-Navigator on NWEA Geometry

Overall Score on the NWEA Geometry Test

In this section we address geometry outcomes using the NWEA Geometry scale. Table 53 provides a summary of the sample we used and the results for the comparison of NWEA Geometry scores for students in *GC+Nav* and control groups. The "Unadjusted" row gives information about all the students in the original sample for whom we have a pretest and posttest. This shows the means and standard deviations as well as counts for students, classes, and teachers in that group. The last two columns provide the effect size, that is, the size of the difference between the means for *GC+Nav* and control in standard deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference as large as, or larger than, the absolute value of the one observed when there truly is no difference. The "Adjusted" row is based on the same sample of students. The means, and therefore the effect size, are adjusted to take into account the student pretest scores; hence, these statistics are adjusted for any chance imbalances on the pretest between the two randomized groups[20].

[20] All statistical equations that follow build on the one used to compute the adjusted effect size.

Table 53. Overview of Sample and Impact of TI-Navigator on NWEA Geometry

	Condition	Means	Standard deviations[a]	No. of students	No. of classes	No. of teachers	Effect size	*p* value[b]	Percentile standing
Un-adjusted effect size	**Control**	-0.06	1.04	199	11	7	0.22	0.83	8.71%
	GC+Nav	0.15	0.88	257	17	7			
Adjusted effect size	**Control**	-0.06	1.04	The same sample is used in both calculations			0.14	0.16	5.57%
	GC+Nav	0.08	0.88						

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the scores of the students in the sample for that row.

[b] The *p* value for the unadjusted effect size is computed using a model that figures in clustering of students in teacher level but does not adjust for any other covariates. The *p* value for the adjusted effect size is computed using a model that controls for clustering and that includes both the pretest and indicators for upper-level units within which the units of randomization are nested (pairs).

[c] Modeling separate intercepts for upper-level units leads to estimates of performance, in the absence of treatment, which are specific to those units. For purposes of display, to set the performance estimate for the control group, we compute the average performance for the sample of control cases used to calculate the adjusted effect size. The estimated treatment effect, which is constrained to be constant across upper-level units, is added to this estimate to show the relative advantage or disadvantage to being in the treatment group.

Figure 13 provides a visual display of results from the row labeled “Adjusted” in Table 53. The bar graphs represent average performance using the metric of NWEA Geometry. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students’ pretest scores and other fixed effects. The overall algebra effect size (in standard deviation units) is .14, which is equivalent to a gain of approximately 6.0 percentile points for the median control group student if the student had received TI-Navigator*.* The *p* value for the *GC+Nav* effect (.16) indicates we should have limited confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The non overlap in these intervals expresses the confidence level noted above.

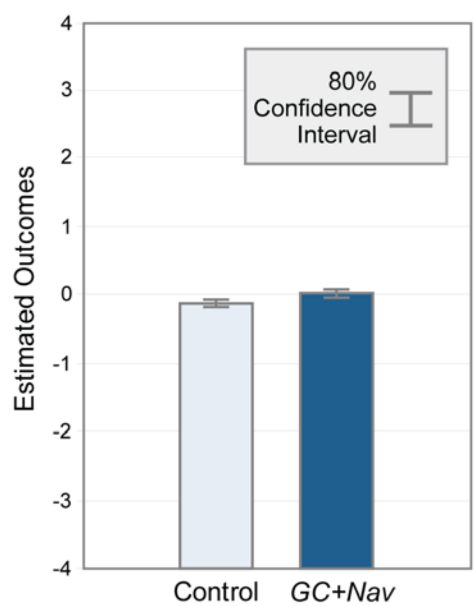


Figure 13. Impact on NWEA Geometry Achievement: Adjusted Means for Control and *GC+Nav*

Moderating Variables

We now report on our examination of the moderating effects of other variables (e.g., performance on pretest, gender, and English proficiency). We provide a separate table of results for each of these moderator analyses. The fixed factor part of each table provides estimates of the effects of interest.

Including Pretest as a Moderator

We first show whether the impact of TI-Navigator is different for students at different levels of prior achievement. At the bottom of the table we give results for technical review—these often consist of what are called random effects estimates. As was described earlier in this report, random effects are added to the statistical equation to account for dependencies in observed scores that happen because students come from the same classes or teachers[21].

Table 54. Impact of TI-Navigator on NWEA Geometry

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a control student with an average pretest	0.05	0.11	5	0.5	.64
Change in outcome for a control student for each unit-increase on the pretest	0.69	0.05	439	14.76	<.01
Effect of for a student with an average pretest	0.14	0.08	5	1.61	.17
Change in the effect of for each unit-increase on the pretest	0.02	0.06	439	0.38	.7
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.01	0.01		0.76	.22
Within-teacher variation	0.31	0.02		14.83	<.01

[a] Although we also modeled differences among some matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

Table 54 shows the estimated impact of TI-Navigator on the performance of students with an average pretest as measured by NWEA Geometry as well as the moderating effect of their prior scores.

[21] In some cases, to account for these dependencies, we model fixed rather than random effects but do not present the individual fixed effects estimates in the table.

The row in the table labeled “Effect of *GC+Nav* for a student with an average pretest” tells us whether TI-Navigator made a difference in the NWEA Geometry scores for a student who has an average score on the pretest. The estimate associated with *GC+Nav* is 0.14. This shows a positive difference associated with *GC+Nav*. The *p* value of 0.17 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 17% of the time when there truly is no effect. Using the criteria outlined earlier in the report, we conclude that we have limited confidence that the true impact is different from zero.

We also estimated the moderating effect of the pretest score on the impact of *GC+Nav* (row 4) to determine whether the intervention was differentially effective for students at different points along the pretest scale. The coefficient associated with the interaction of pretest with treatment is 0.02, which shows a small increase in the treatment effect with each one-unit increase on the pretest. The *p* value of .70 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 70% of the time when in fact there is zero differential impact[22]. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true differential impact is different from zero.

As a visual representation of the result described in Table 54, we present a scatterplot in Figure 14, which shows student performance at the end of the year in Geometry as measured by the NWEA Geometry test, against their performance on NWEA in the fall.. This graph shows where each student started in terms of his or her pretest score (horizontal x-axis) and his or her outcome score[23]. We remind the reader that pretest scores have been z-transformed within test-type. Each point plots one student’s post-intervention score against his or her pre-intervention score. The darker points represent *GC+Nav* students; the lighter points, control students.

The two lines are the estimated values on the posttest for students in the *GC+Nav* and control conditions. We see very little difference in impact across the prior score scale. [24]

[22] In the model used, intercepts are modeled as random at the student and teacher levels and as fixed at the pair level; however, slopes are not modeled as random; the interaction of pretest with treatment and the corresponding *p* value do not reflect uncertainty due to a potential re-sampling of teachers.

[23] The value for each point along the horizontal scale represents the distance from the origin, in standard deviation units.

[24] Pairs were modeled as a fixed factor, resulting in a separate intercept estimate for each pair.

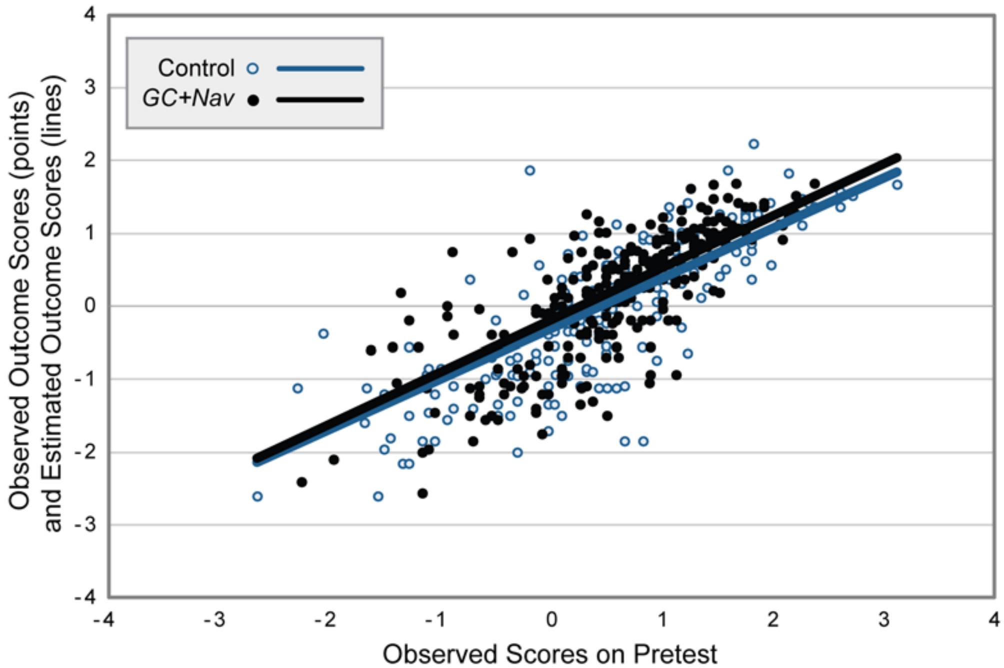


Figure 14. Comparison of Estimated and Actual Outcomes for *GC+Nav* and Control Group Students (NWEA Geometry Achievement)

Including Gender as a Moderator

In addition to looking at the main effect of TI-Navigator, we estimated the interactions of *GC+Nav* with the pretest scores and gender of the students. In particular, we were interested in whether the condition's effect was different for girls and boys. Table 55 shows the moderating effect of gender on students' performance on NWEA Geometry.

The advantage of being in the *GC+Nav* condition is greater for boys than it is for girls. The *p* value of .21 means we have no confidence that the actual differential impact is different from zero[25].

[25] When we limit the analysis to only those students who took MAP posttest, the p-value for the interaction drops to .08, with *GC+Nav* being more beneficial for males.

Table 55. Moderating Effect of Gender on NWEA Geometry

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the control girl with an average pretest	0.08	0.11	5	0.73	.50
Change in outcome for each unit-increase on the pretest	0.7	0.03	438	20.68	<.01
Control group difference (boys minus girls) in the outcome	-0.06	0.08	438	-0.7	.49
Effect of *GC+Nav* for girls	0.08	0.1	5	0.79	.46
Net difference (boys minus girls) in the effect of *GC+Nav*	0.14	0.11	438	1.25	.21.
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.01	0.01		0.69	.24
Within-teacher variation	0.31	0.02		14.81	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 456 students we have information on gender, we removed 1 because it was influential points or outliers. Hence, 455 students were used in the moderator model.

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

Including English Proficiency as a Moderator

We were also interested in the moderating effect of student English proficiency. In particular, we wanted to know whether TI-Navigator was differentially effective for English proficient students and for English learners. Table 56 shows the results for the moderating effect of student English proficiency.

The advantage of being in the *GC+Nav* condition is negligibly greater for English proficient students. The *p* value of .69 means we have no confidence that the actual differential impact is different from zero[26].

[26] When we limit the analysis to only those students who took MAP posttest, the *p* value for the interaction drops to .10, with *GC+Nav* being more beneficial for English proficient students.

Table 56. Moderating Effect of English Proficiency on NWEA Geometry

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the Non-English proficient control with an average pretest	-0.02	0.13	5	-0.13	.9
Change in outcome for each unit-increase on the pretest	0.68	0.04	437	19.24	<.01
Control group difference (English proficient minus not proficient) in the outcome	0.12	0.13	437	0.91	.36
Effect of *GC+Nav* for Non-English proficient student	0.1	0.12	5	0.81	.46
Net difference (English proficient minus not proficient) in the effect of *GC+Nav*	0.06	0.14	437	0.4	.69
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.01	0.01		0.49	.31
Within-teacher variation	0.31	0.02		14.78	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 456 students we have information on English proficiency level; we removed 1 because it was influential points or outliers. Hence, 455 students were used in this moderator model.

Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

The estimate of no difference between English-speakers and English-learners among controls is due to the inclusion of pretest in the model, which makes the 'English-speaker effect be 'net of' pretest. When we exclude pretest from the model, the average for English-speakers is larger than the average for non-English speakers.

Impact of TI-Navigator on CST Geometry

Overall Score on the CST Geometry Test

We now address geometry outcomes using the CST Geometry scale. Table 57 provides a summary of the sample we used and the results for the comparison of CST Geometry scores for students in *GC+Nav* and control groups. The "Unadjusted" row gives information about all the students in the original sample for whom we have a pretest and posttest. This shows the means and standard deviations as well as counts for students, classes, and teachers in that group. The last two columns provide the effect size, that is, the size of the difference between the means for *GC+Nav* and control in standard deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference as large as, or larger than, the absolute value of the one observed when there truly is no difference. The "Adjusted" row is based on the same sample of students. The means, and therefore the effect size, are adjusted to take into account the student pretest scores; hence, these statistics are adjusted for any chance imbalances on the pretest between the two randomized groups. [27]

Table 57. Overview of Sample and Impact of TI-Navigator on CST Geometry

	Condition	Means	Standard deviations[a]	No. of students	No. of classes	No. of teachers	Effect size	*p* value[b]	Percentile standing
Un-adjusted effect size	Control	298.01	62.57	197	11	7	0.12	0.90	4.78%
	GC+Nav	305.50	58.33	277	19	7			
Adjusted effect size	Control	298.01	62.57	The same sample is used in both calculations			0.02	0.82	0.80%
	GC+Nav	299.37[c]	58.33						

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the scores of the students in the sample for that row.

[b] The *p* value for the unadjusted effect size is computed using a model that figures in clustering of students in teacher level but does not adjust for any other covariates. The *p* value for the adjusted effect size is computed using a model that controls for clustering and that includes both the pretest and, where relevant, indicators for upper-level units within which the units of randomization are nested.

[c] Modeling separate intercepts for upper-level units leads to estimates of performance, in the absence of treatment, which are specific to those units. For purposes of display, to set the performance estimate for the control group, we compute the average performance for the sample of control cases used to calculate the adjusted effect size. The estimated treatment effect, which is constrained to be constant across upper-level units, is added to this estimate to show the relative advantage or disadvantage to being in the treatment group.

[27] The goal of modeling the pretest and pair effects is to increase the precision of the estimate of the program effect by accounting for variation due to these factors. These effect estimates are usually carried forward to all subsequent statistical equations that we use. (For this scale, we modeled a set of pairs that both permitted estimation of the random intercept at the teacher-level and maximized model-fit.)

Figure 15 provides a visual display of results from the row labeled "Adjusted" in Table 57. The bar graphs represent average performance using the metric of CST Geometry. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students' pretest scores and other fixed effects. The overall geometry effect size (in standard deviation units) is.02, which is equivalent to a gain of .8 of a percentile point for the median control group student, if the student had received TI-Navigator. The high *p* value for the treatment effect (.82) indicates we should have no confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The overlap in these intervals further indicates that any difference we see is easily due to chance.

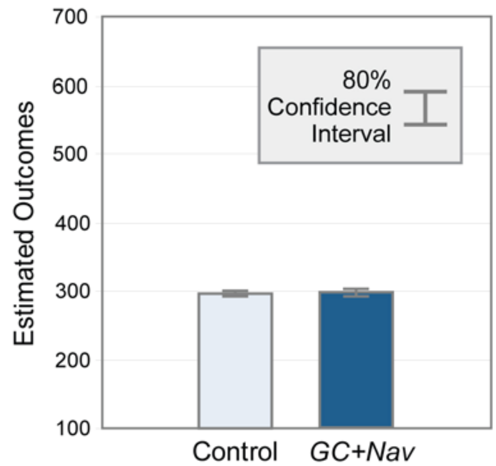


Figure 15. Impact on CST Geometry Achievement: Adjusted Means for Control and *GC+Nav*

Moderating Variables

We now report on our examination of the moderating effects of other variables (e.g., performance on pretest, gender, and English proficiency). We provide a separate table of results for each of these moderator analyses. The fixed factor part of each table provides estimates of the effects of interest.

Including Pretest as a Moderator

We first show whether the impact of TI-Navigator is different for students at different levels of prior achievement. At the bottom of the table we give results for technical review—these often consist of what are called random effects estimates. As was described earlier in this report, random effects are added to the statistical equation to account for dependencies in observed scores that happen because students come from the same classes or teachers[28].

[28] In some cases, to account for these dependencies, we model fixed rather than random effects but do not present the individual fixed effects estimates in the table.

Table 58. Impact of TI-Navigator on CST Geometry

Fixed effects[a]	Estimate	Standard error	DF	t value	*p* value
Outcome for the control student with an average pretest	296.85	6.76	5	43.95	<.01
Change in outcome for the control student for each unit-increase on the pretest	44.81	3.41	456	13.15	<.01
Effect of *GC+Nav* for a student with an average pretest	1.31	5.82	5	0.23	.83
Change in the effect of *GC+Nav* for each unit-increase on the pretest	2.93	4.53	456	0.65	.52
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	45.38	59.8		0.76	.22
Within-teacher variation	1513.93	100.3		15.09	<.01

[a] Although we also modeled differences among some matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

Table 58 shows the estimated impact of TI-Navigator on the performance of students with an average pretest score as measured by CST Geometry as well as the moderating effect of their prior scores.

The row in the table labeled “Effect of *GC+Nav* for a student with an average pretest” tells us whether TI-Navigator made a difference in the CST Geometry scores for a student who has an average score on the pretest. The estimate associated with *GC+Nav* is 1.31. This shows a small positive effect associated with *GC+Nav*. The *p* value of 0.83 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 83% of the time when there truly is no effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero.

We also estimated the moderating effect of the pretest score on the impact of *GC+Nav* (row 4) to determine whether the intervention was differentially effective for students at different points along the pretest scale. The coefficient associated with the interaction of pretest with treatment is 2.93, which shows a small increase in the treatment effect with each one-unit increase on the pretest. The *p* value of .52 indicates that we can expect to see a difference, as large as or larger than the absolute value of the estimate, 52% of the time when in fact there is zero differential impact[29]. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true differential impact is different from zero.

As a visual representation of the result described in Table 58, we present a scatterplot in Figure 16, which shows student performance at the end of the year in Geometry as measured by the CST Geometry test, against their performance on NWEA in the fall. This graph shows where each student started in terms of his or her pretest score (horizontal x-axis) and his or her outcome score[30]. We remind the reader that pretest scores have been z-transformed within test-type. Each point plots one student's post-intervention score against his or her pre-intervention score. The darker points represent *GC+Nav* students; the lighter points, control students.

The two lines are the estimated values on the posttest for students in the *GC+Nav* and control conditions. We see very little difference in impact across the prior score scale. Consistent with the results described above, we observe that TI-Navigator and the programs used in the control classrooms were nearly equally effective as measured by CST Geometry[31] [32].

---

[29] In the model used, intercepts are modeled as random at the student and teacher levels and as fixed at the pair level; however, slopes are not modeled as random; the interaction of pretest with treatment and the corresponding *p* value do not reflect uncertainty due to a potential re-sampling of teachers.

[30] The value along the horizontal scale for each point represents the distance from the origin, in standard deviation units.

[31] Pairs were modeled as a fixed factor, resulting in a separate intercept estimate for each pair. To fix the vertical location of the prediction lines, we selected the median estimate from among the pair estimates for the intercept.

[32] We observe two features of the scatterplot. First, there appears to be heteroscedasticity - variance in the posttest is the not the same across the scale of the pretest. To address this we use the 'unstructured' option in SAS which leaves the fitted covariance matrix for observations unconstrained. Second, we observe that there is a slight curvilinear trend in the scatter – there seems to be a non-linear rise in the trend of the scatter as you move to the right. This is probably due to characteristics of the CST scale. Tests of skew and kurtosis in the outcome indicated that, according to conventional standards, there was no need to transform posttest scores. Nonetheless, we were concerned that that the estimate of the interaction effect may be driven by the fact that we fit straight-line effects on a curvilinear scatter. To check this possibility we also used a model that included a quadratic term for the pretest as well as a term for the interaction between the pretest-squared and treatment. The results were essentially unchanged.

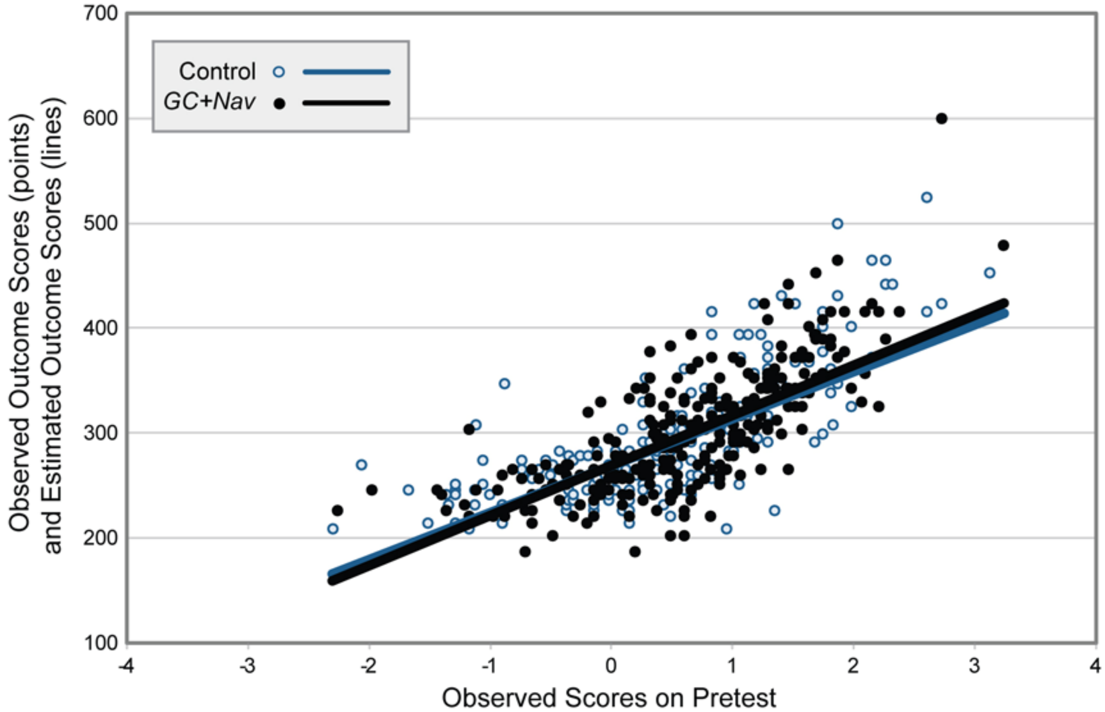


Figure 16. Comparison of Estimated and Actual Outcomes for *GC+Nav* and Control Group Students (CST Geometry Achievement)

Including Gender as a Moderator

In addition to looking at the main effect of TI-Navigator, and the differential effect of pretest, we estimated the interaction of *GC+Nav* with gender. In particular, we were interested in whether the condition's effect was different for girls and boys. Table 59 shows the moderating effect of gender on students' performance on CST Geometry.

We see that there is essentially no difference in the treatment effect between boys and girls. The *p* value for the moderating effect is .27, which gives us no confidence that the true effect is different from zero.

Table 59. Moderating Effect of Gender on CST Geometry

Fixed effects [a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the control girl with an average pretest	304.32	7.3	5	41.66	<.01
Change in outcome for each unit-increase on the pretest	46.5	2.34	455	19.91	<.01
Control group difference (boys minus girls) in the outcome	-13.3	5.58	455	-2.39	.02
Effect of *GC+Nav* for girls	-3.01	6.79	5	-0.44	.68
Net difference (boys minus girls) in the effect of *GC+Nav*	8.13	7.34	455	1.11	.27
Random effects [b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	44.77	58.21		0.77	.22
Within-teacher variation	1496.23	99.24		15.08	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 474 students we have information on gender, we removed 2 because they were influential points or outliers. Hence, 472 students were used in the impact model.

[a] Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

[b] Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers..

Including English Proficiency as a Moderator

We were also interested in the moderating effect of student English proficiency. In particular, we wanted to know whether TI-Navigator was differentially effective for English proficient students and for English learners. Table 60 shows the results for the moderating effect of student English proficiency.

We see that there is essentially no difference in the treatment effect between English proficient and non-proficient students. The *p* value for the moderating effect is .57, which gives us no confidence that the true effect is different from zero.

Table 60. Moderating Effect of English Proficiency on CST Geometry

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for the Non-English proficient control with an average pretest	303.49	9.5	5	31.94	<.01
Change in outcome for each unit-increase on the pretest	46.94	2.43	453	19.32	<.01
Control group difference (English proficient minus not proficient) in the outcome	-8.05	8.95	453	-0.9	.37
Effect of *GC+Nav* for Non-English proficient student	-2.87	9.4	5	-0.3	.77
Net difference (English proficient minus not proficient) in the effect of *GC+Nav*	5.75	10.1	453	0.57	.57
Random effects	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	43.12	57.6		0.75	.23
Within-teacher variation	1519.3	100.9		15.06	<.01

Notes. Rows 3 through 5 provide the estimated average effect regardless of the students' prior test scores. Row 5 = row 3 minus row 4. Of the 472 students we used to calculate the adjusted effect size, we removed 2 because they were influential points or outliers. Hence, 470 students were used in the impact model.

Although we also modeled differences among matched pairs by estimating separate fixed effects, we do not exhibit these estimates in the table.

Teachers were modeled as a random factor. Teacher mean achievement represents the variation among teacher-averages of student outcomes net of the treatment effect. Within-teacher variation represents the average variability in student outcomes within teachers.

The estimate of no difference between English-speakers and English-learners among controls is due to the inclusion of pretest in the model, which makes the 'English-speaker effect' be 'net of' pretest. When we exclude pretest from the model, the average for English-speakers is larger than the average for non-English speakers.

Discussion

Our experiment addressing the effectiveness of TI-Navigator was tightly defined, since the comparison group was not typical of many effectiveness studies, but instead consisted of classrooms well equipped with graphing calculators and display systems, and staffed by teachers trained in their use. Beyond the basic impact of graphing calculators, we were investigating whether there was an additional impact of providing teachers with a graphing calculator networking system (and training) that afforded greater interactivity in the classroom. This was the second year of a two-year randomized control trial. In the first year, we used a matched pair design to randomly assign 44 teachers to use graphing calculators with their existing math curriculum or to conduct "business as usual" in the classroom. In this second year, teachers kept their random assignments and the original graphing calculator group received TI-Navigator while the original control group received graphing calculators.

Although for the most part, the experiment could not discern an impact of TI-Navigator, there was some evidence of an impact for Geometry classes. In Geometry classes we found a modest effect (the equivalent of about 6 percentile points) when we measured achievement using the NWEA End of Course Geometry test, but this impact was not reflected in CST Geometry test scores. In Algebra classes, we found no overall difference as a result of providing the equipment and training for TI-Navigator. There was, however, some evidence of a small negative impact for students scoring "below basic" on the CST Algebra test and, holding pretest score constant, there was a small negative impact for English proficient students on the same test. The results of the NWEA End of Course Algebra I test did not reflect those same results.

An effectiveness study of this design does not consider differences in implementation as part of the impact calculations; nonetheless, these differences must be considered in interpreting the results of those calculations. Our extensive surveys and observations make clear that this implementation was not a fair test of what difference TI-Navigator might make if used more extensively. Of the 19 teachers originally assigned to the treatment group, about half did not use the system at all for instruction. Of the remaining nine teachers, only three teachers could be considered "Comprehensive-Implementers." Of those three teachers, only one teacher used TI-Navigator on a daily basis. Technical glitches deterred many teachers from using the system after previous failed attempts at use. Overall use of the technology may have been constrained by the fact that California prohibits calculator use on the state tests.

Our results also must be qualified by the fact that while finding differences on one test, we did not find differences on the other test. In Algebra, differences were found for CST Algebra I but not for NWEA End of Course Algebra I. In Geometry our positive result for achievement measured by NWEA Geometry was not found for CST Geometry. In addition, inconsistencies among the ALT and MAP versions of the NWEA tests used raises concerns about their interpretation. The significant amount of attrition, both at the teacher and student levels, although not believed to be associated with the program being tested, does raise issues about generalizability. For example, after the randomization, some teachers who taught algebra and geometry classes selected which of their classes were in the study. Additionally, it is clear that, in both experimental conditions, lower scoring students were significantly more likely to not have posttests, indicating that the results are not applicable to the lowest scoring students in these districts.

Overall, we did find that the TI-Navigator had an impact on the average number of minutes the technology was used. The teachers with TI-Navigator reported using the technology about 15 minutes more per week per class period than teachers without. Future exploratory analyses may prove useful in suggesting whether extent of usage can account for student outcomes. In particular, since TI-Navigator resulted in greater technology use, examining the correlation between technology use and achievement may suggest a mechanism by which TI-Navigator could be effective. It is interesting to note that the teachers coded as having a "Comprehensive-Implementation" were Geometry teachers who had had an extra semester of experience in using the networking system. Future studies of TI-Navigator will benefit from greater support for implementation. We also recommend continuing to include Geometry in the topics to which TI-Navigator is applied since the positive result found in this experiment should be replicated.

We designed the experiment described in this report to provide useful information to the participating school districts. Because we were testing a specific implementation of TI-Navigator in a particular setting, we caution readers about generalizing the results to districts with different populations, resources, and other relevant conditions. Although our results cannot be used as definitive evidence of the value of TI-Navigator, the areas of positive findings lead us to recommend that schools planning to implement TI-Navigator provide adequate support, both technically and educationally, while rolling out implementation in a manner that allows for continued tracking of student achievement gains.

References

Heller, J., Curtis, D. Jaffe, R., and Verboncoeur, C. (2006). *Impact of Hand held Graphing Calculator Use on Student Achievement in Beginning Algebra*. Retrieved December 20, 2006 from: *http://education.ti.com/sites/US/downloads/pdf/heller_grcalcreport_2005.pdf*

Miller, G.I., Jaciw, A. Hoshiko, B., & Vu, M., & Wei, X. (2007, August). *The Effect of TI-84 Graphing Calculators on Algebra I and Geometry Outcomes: A report of randomized experiments in two school districts.* Palo Alto, CA: Empirical Education Inc.

National Center for Education Statistics. (2005). *District Detail for East Side Union High*. Retrieved December 14, 2007 from: *http://www.nces.ed.gov/ccd/districtsearch/district_detail.asp?Search=2&DistrictID=0611820&ID2=0611820&details=*

National Center for Education Statistics. (2005). *District Detail for San Diego Unified*. Retrieved December 14, 2007 from: *http://www.nces.ed.gov/ccd/districtsearch/district_detail.asp?Search=2&DistrictID=0634320&ID2=0634320&details*

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs For Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

United States Census Bureau. (2000) *San Diego city, California – Fact Sheet.* Retrieved December 14, 2007 from: http://factfinder.census.gov/servlet/SAFFFacts?_event=ChangeGeoContext&geo_id=16000US0666000&_geoContext=01000US%7C04000US06%7C16000US0668000&_street=&_county=san+diego&_cityTown=san+diego&_state=04000US06&_zip=&_lang=en&_sse=on&ActiveGeoDiv=geoSelect&_useEV=&pctxt=fph&pgsl=010&_submenuId=factsheet_1&ds_name=DEC_2000_SAFF&_ci_nbr=null&qr_name=null&reg=null%3Anull&_keyword=&_industry=

United States Census Bureau. (2000) San Jose city, California – Fact Sheet. Retrieved December 14, 2007 from: *http://factfinder.census.gov/servlet/SAFFFacts?_event=&geo_id=16000US0668000&_geoContext=01000US%7C04000US06%7C16000US0668000&_street=&_county=san+jose&_cityTown=san+jose&_state=04000US06&_zip=&_lang=en&_sse=on&ActiveGeoDiv=&_useEV=&pctxt=fph&pgsl=160&_submenuId=factsheet_1&ds_name=ACS_2006_SAFF&_ci_nbr=null&qr_name=null&reg=&_keyword=&_industry*