Spatial Temporal Mathematics at Scale:
An Innovative and Fully Developed Paradigm to Boost Math Achievement Among All Learners

Teomara Rutherford[1], Melissa Kibrick[1], Margaret Burchinal[1], Lindsey Richland[1], AnneMarie Conley[1], Keara Osborne[1], Stephanie Schneider[2], Lauren Duran[2], Andrew Coulson[3], Fran Antenore[3], Abby Daniels[3], and Michael E. Martinez[1]

Paper presented at the annual convention of the American Educational Research Association
Denver CO, May 2010

The advancement of science and technology is vital to sustained economic success in the United States. The next generation of STEM professionals must be prepared to engage the world's most pressing problems. Education, especially in the science, technology, engineering and math (STEM) disciplines, is vital to this endeavor. Yet the U.S. public education system is not adequately developing the intellectual readiness needed to sustain the nation's economy (NAS, 2007). Students graduating from U.S. high schools are seriously under-prepared to enter STEM fields, which require academic proficiency in mathematics and science, as well as proficiency in problem solving and reasoning.

International comparisons show that the performance of American students is far below optimal. In the most recent PISA assessment, for example, top performing students in the U.S. ranked well behind students in the highest achieving countries (PISA, 2007). Performance lagged especially in mathematics, with U.S. proficiency falling well below the average for industrialized countries. In the U.S., one-third of 4th graders and one-fifth of 8th graders lack the competence to perform basic mathematical computations, with low-income children performing below middle-income children (NAEP, 2005). Achievement gaps are also evident when comparing African American and Hispanic students with white or Asian American students (Levitt & Fryer, 2004) and when comparing girls and boys (Nation's Report Card, 2009). The United States

---

[1] University of California, Irvine

[2] Orange County Department of Education

[3] MIND Research Institute

remains one of the few primary industrialized nations to show a significant gender difference in mathematics performance as early as fourth grade (Ginsburg, Cooke, Leinwand, Noell, & Pollock, 2005, p. 19). Compounding these disappointing achievement results, one-third of 9th grade students in the U.S. do not graduate from high school (Barton, 2005), a shortcoming that inevitably impacts the U.S. economy and society (ACTE, 2007).

The No Child Left Behind Act (NCLB) of 2002 was intended to address inadequate achievement in mathematics and other subject domains, as well as discrepancies in proficiency between subgroups. According to the logic of NCLB, all students must be "proficient" in math and reading by 2014, a timeline that the Obama administration now regards as unrealistic (United States Department of Education, 2010). NCLB legislation prescribed that schools and school districts must raise the proficiency of at-risk student subgroups in order to meet targets. Unfortunately, many factors conspire against attainment of these proficiency goals. One important factor is that many teachers lack the necessary content knowledge: A third of secondary school math teachers teach out of their field, and teachers at low-income schools are more likely to be teaching out of their field than teachers at affluent schools (Ingersoll, 2002). According to the 2008 report issued by the National Mathematics Advisory Panel, "Research on the relationship between teachers' mathematical knowledge and students' achievement confirms the importance of teachers' content knowledge" (NMAP, 2008, p. xxi). The Panel articulated the desperate need to improve teachers' mathematical knowledge, particularly at the elementary and middle school levels.

Educational research has the potential to guide effective solutions to these longstanding dilemmas. Unfortunately, the search for effective approaches to teaching and learning mathematics is limited by the existing research base. The research literature comprises few rigorous evaluations of mathematics curricula or instructional practices, especially evaluations that permit inferences about causal associations that are separable from confounds. Among 237 studies that examined the efficacy of particular interventions in mathematics education, only two studies met strict clinical trial criteria or random assignment to conditions. A mere seven studies met slightly lower standards (What Works Clearinghouse, 2008).

This paper describes a project designed to elevate student math achievement through a large-scale randomized field trial of Spatial-Temporal Math (ST Math), a large suite of interactive mathematics software. ST Math provides individualized delivery of a standards-based mathematics curriculum by capitalizing on the fact that many fundamental math concepts—fractions, proportional reasoning, symmetry, and arithmetic operations—can be presented as pictorial images. In ST Math software games, students solve problems by responding to problems presented within an image-based medium. By adherence to the “gold standard” of randomization (Shadish, Cook, & Campbell, 2002), this study aims to establish a causal relationship between student participation in the ST Math program and increases in positive educational outcomes, including California Standards Test (CST) scores in mathematics, narrower measures of math achievement and ability, and student motivation.

In this paper, we present the construct of spatial cognition, the ST Math approach, and why we believe some student groups may particularly benefit from the spatial-temporal approach to learning mathematical skills and concepts. We then present preliminary results from treatment-control comparisons on the California Standard Test after one year of implementation, as well as preview anticipated study directions.

Spatial Cognition & ST Math

Information-processing models of spatial cognition date back to Kosslyn’s model of spatial cognition as akin to a computer monitor (Kosslyn, 1980). Psychometric models of spatial ability extend back much further, dating at least to the research of Thurstone in the 1930s. Spatial cognition is a widely-held cognitive ability that manifests as the capacity to mentally hold and manipulate a two- or three-dimensional image, often in the service of solving a problem (Dennis & Tapsfield, 1996). In addition to mental manipulation, spatial cognition also entails the ability to perceive patterns and forms amidst visual noise, and to compare figures and shapes (Carroll, p. 309). Spatial abilities have been linked to expertise in STEM fields, including engineering and physics; some of science’s most brilliant minds have referenced their dependence on spatial thinking in problem solving (Martinez et al., 2008; Shaw, 2000; Sorby, 2009). This may be because many scientific phenomena can be represented as “mental models”—three-dimensional cognitive depictions that can be “run” to indicate temporal dynamics and to make predictions. The propensity of boys to think spatially has even been proposed as one explanation for the science achievement gap between boys and girls (AAUW, 2010, p.52).

Recognizing the importance of spatial cognition, several organizations concerned with STEM education have called for programs that draw upon spatial cognition to advance learning outcomes (National Council of Teachers of Mathematics, 2000; National Research Council, Committee on Support for Thinking Spatially, 2006). This represents a rather significant departure from traditional pedagogy: much domain content in today's classrooms is taught through verbal-analytic methods that emphasize language and symbols, rather than images (Grandin, Peterson, & Shaw, 1998; Shaw, 2000, Sorby, 2009). Broad acceptance of a rationale for teaching spatial thinking has yet to be achieved among the educational community, as manifest by the absence of formal standards on the teaching of spatial thinking (Committee on Support for Thinking Spatially, 2006, p. 6). This lack of attention may have serious implications for students. International tests of math and science indicate particular weaknesses among U.S. students in measurement and geometry, two areas related to spatial representations (Ginsburg, Cooke, Leinwand, Noell, & Pollock, 2005, p. 16). Some scientists and education researchers worry that the U.S. may not be able to meet the need for proficient spatial thinkers in a modern, global economy unless students across all stages of schooling are given the appropriate tools and experiences (Committee on Support for Thinking Spatially, 2006; Shaw, 2000; Sorby, 2007). ST Math may meet an important need by teaching spatial thinking in the context of mathematics and, significantly, mathematics in the context of dynamic spatial representations.

Optimism about the pedagogical value of ST Math is supported by research to date. Earlier quasi-experimental and smaller-scale experimental studies showed positive effects from the training of spatial thinking on mathematics outcomes (Martinez et al., 2008; Peterson et al.; 2004; Graziano, Peterson & Shaw, 1998). Comparable effects of spatial training have also been found in college populations (Sorby, 2007; Sorby, 2009). ST Math initially minimizes the use of mathematical symbols, terminology, and language in general. This strategy might make mathematics concepts more widely accessible; conventional language- and symbol-based abstractions may present, for many students, unnecessary complications to initial learning of math concepts. Moreover, heavy use of symbols and technical language may impede not only learning, but also the enjoyment and motivation that can spring from mathematical insight. Instead of traditional symbolic expressions, the spatial temporal approach employs intentionally simple, yet engaging, dynamic shapes as representations in the mathematical puzzles to be solved. With potential distractions removed or minimized, the learner encounters a robustly intuitive visual problem-solving workspace within which to build conceptual understanding and problem-solving skills.

ST Math is delivered through a 1:1, interactive, animated learning environment wherein students work at their own pace. The animation of the visualizations is important to providing instructive feedback to both correct and incorrect solutions. Correct solutions are animated to show why, heuristically, they are correct. Likewise, incorrect solutions are interactively animated to show why they fall short, and often to indicate how the response differs from an ideal solution. Feedback of this sort has been shown to be a highly desirable characteristic of both games and instruction, giving the learner valuable information to guide their progress toward the goal of self-regulated learning (Garris, Ahlers, & Driskell, 2002; Metcalfe & Kornell, 2007). ST Math provides a very high frequency (on average, twice per minute) of corrective feedback, and allows for the gradual extrapolation of mathematics principles within lessons, all the while building the student's self-confidence and motivation.

A second motivating aspect of ST Math is an academic exploitation of the popular video-game metaphor. A suite of game-like exercises engages and motivates students to solve mathematics problems to steadily advance. A simple and yet somewhat rare feature of educational software is the very carefully calibrated and incrementally scaffolded difficulty levels built into ST Math. The initial low difficulty level conveys immediate success. Once that level is "won," the reward is a new and slightly more difficult level with a small extrapolation of math principles, such as extension to larger quantities. Games proceed in this way by stages, often leading to quite challenging, multi-step problem solving. A video describing the differences between the spatial temporal approach to math and conventional math can be viewed at the MIND Research website, www.mindresearch.net/video/demo.html.

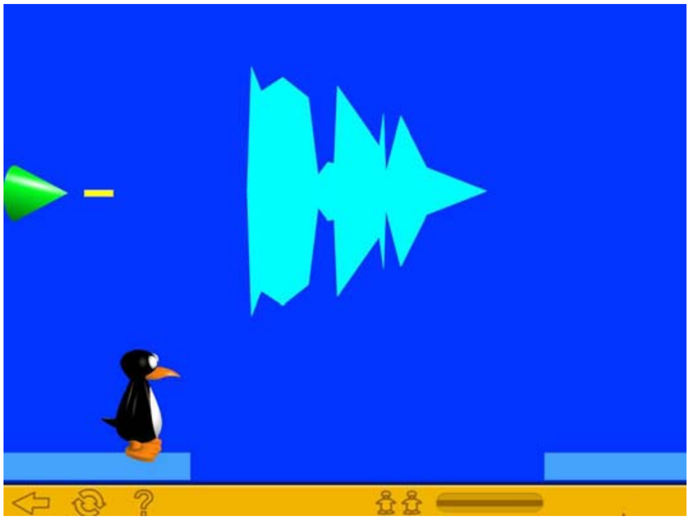
Examples of this scaffolding, an illustrative span of content, and ranges of difficulty can be seen in the follow examples:

Grade 2

Module 6: Geometry & Measurement

Game 3: Ice Caves

Figure 1.



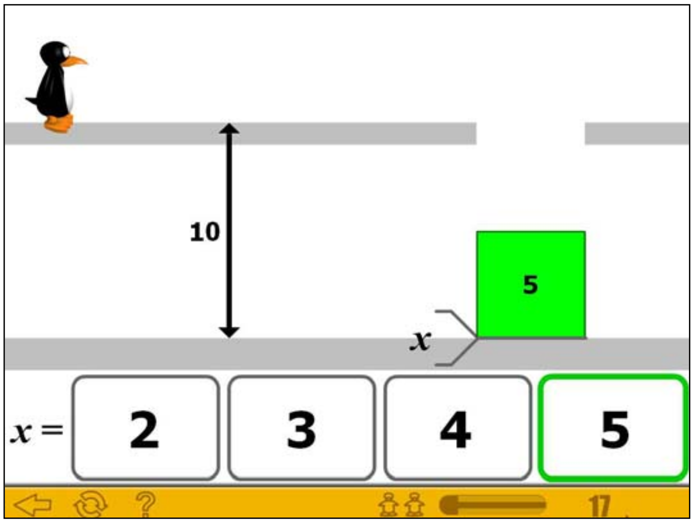
Difficulty Level 1. At this first level, there is only one green launcher. Once clicked, a yellow dash splits the symmetrical shape in half along its axis of symmetry. When the task is complete, the blue gap at the bottom of the screen is filled and JiJi, the penguin, is able to exit the right side of the screen.

Figure 2.

Difficulty Level 6. At this sixth of eight difficulty levels, the student has 12 choices of green launcher to clear 16 shapes.

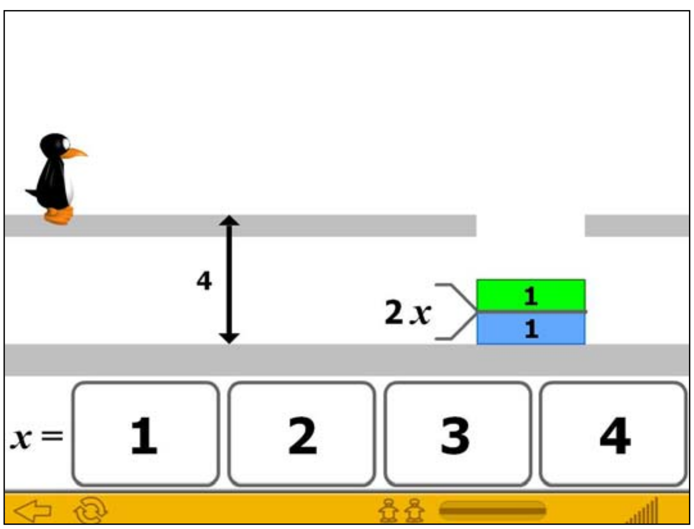
Move #2 is in progress: The bottom left green launcher has just been clicked, and the yellow dash will split all 3 symmetrical shapes above the launcher (the first is in the process of being split, with two remaining above it). Planning is required to select the correct sequence of moves. A random sequence will probably not solve the puzzle. However, more than one sequence will solve the puzzle, allowing for multiple solutions.

Figure 3.



<u>Difficulty Level 1</u>. Students must decide what height of block needs to be added to the existing green block of height 5 in order to make a "bridge" of height 10. This will complete the grey line so that JiJi, the penguin, can exit to the right side of the screen. This puzzle, which can be solved with visual reasoning, represents the algebraic equation: $10 = x + 5$*, solve for x.*

Figure 4.



<u>Difficulty Level 8</u>. Difficulty has gradually increased to extrapolate to a coefficient in front of the variable. This puzzle can be solved with the assistance of visual reasoning. It represents the algebraic equation:

$4 = 1 + 2x + 1$*, solve for x.*

So far, our research data indicate that the spatial temporal approach fosters strong conceptual understanding of elementary school mathematics. However, in addition to acquiring conceptual knowledge of mathematics, students must become adept at procedural and computational skills (Kilpatrick et al., 2001). But there is a preferred order: Arithmetic skills are most effectively learned and retained if students first understand the *meaning behind* the algorithmic procedures (Brownell & Moser, 1949; Gray, 1965). For this reason, ST Math provides extensive practice for procedural and computational skills. This skill-building component is presented only *after* the student exhibits a conceptual understanding of the topic.

Do Some Students Especially Benefit from ST Math?

We shift our focus now to consider whether some students may especially benefit from a spatial temporal approach to mathematics. ST Math is intended in part to meet the needs and preferences of learners who are underserved by traditional curricula. In The Nation's Report Card, The National Center for Education Statistics (2009) notes that gaps persist in mathematics achievement between native speakers and English Language Learners (ELL), between boys and girls, and between students of different income groups. One very engaging possibility is that these persistent group differences can be redressed in part by presenting mathematical concepts and problems through spatial-temporal representations.

In this study, we consider the categories of gifted students, English Language Learners, girls, and special education students. Spurring the development of ST software was Shaw's (2000) belief in an "innate" spatial-temporal ability—that the human brain is wired to support spatial-temporal cognition and that high levels of this ability are common in scientific thinkers, ranging from Albert Einstein to the autistic savant Temple Grandin. If such abilities are an "innate" characteristic of human brain structure and cognition, then they are resources for all learners. Indeed, Shaw and his colleagues found that young students from disadvantaged backgrounds often displayed strikingly high levels of spatial-temporal ability (Peterson et al., 2004). Traditional schooling, which typically emphasizes verbal-analytic reasoning and methods, may have poorly served spatial geniuses like Einstein and Grandin. Likewise, the pedagogical neglect of spatial reasoning may have resulted in considerable untapped potential among learners now and in the past. Such an imbalance may have privileged some students while disadvantaging others.

English Language Learners. As previously noted, traditional methods of teaching mathematics involve conveying concepts predominantly using language and abstract symbols. Students whose native language is different from the language used for instruction in the classroom may experience particular difficulty. English language learners (ELLs) often have difficulty learning mathematics (Gandara, 2000). Somewhat counterintuitively, ELL students often perform comparatively worse on the mathematics portions of exams than on the language arts portions (Jepson & de Alth, 2005). This situation warrants serious attention, given that approximately 11 percent of students nationwide are ELLs (NCES, 2005). As further illustration of the difficulties faced by ELLs, consider the findings of Wright & Li (2008) who performed a qualitative analysis of the experiences of Cambodian students in a Texas fifth grade classroom. They found that Cambodian students had little opportunity to practice the math skills demanded by the state's standardized tests, resulting in very low scores. They noted that language difficulties, particularly differences between Cambodian and American mathematical notation, along with comparatively poor prior math instruction, contributed to poor performance (Wright & Li, 2008). While these findings may not be generalizable to all immigrants or ELLs, it seems likely that many second-language learners across the country face similar challenges.

ST Math may allow ELLs to master mathematical concepts without simultaneously having to master English-related peculiarities of math learning, while providing a scaffolded introduction to math symbols and language once a conceptual foundation is established. It provides a standards-based grade-level curriculum to non-English speakers, better preparing them for grade-level assessments, like the CSTs, setting a solid foundation on which future math learning can be built.

Gender differences. Although women are increasingly prominent in some traditionally male fields, they remain underrepresented in mathematics and the physical sciences, especially at the doctoral level (AAUW, 2010). This discrepancy may be linked to gender differences in spatial ability. Psychometric research has often suggested that, on average, girls have lower level spatial ability than boys (AAUW, 2010; Linn & Peterson, 1985). However, viewing spatial ability as an insurmountable barrier to women in sciences discounts the potential of significant gains in spatial skill through structured practice. Sorby (2007; 2009), for example, reported gains from a spatial skills training program for first-year engineering students at Michigan Technology University. The gains were especially pronounced among female students, as more females initially exhibited lower spatial scores, but along with their male classmates made gains in tests of spatial cognition and, tellingly, course grades and retention rates (Sorby, 2007; 2009).

Giving both boys and girls the tools to build their spatial ability at an early age with ST Math or similar training programs may help to close the gender-related achievement gap in some STEM disciplines. If spatial representations and problem solving are important to higher math and sciences, bringing girls' skills on par with those of the boys will contribute to leveling the playing field.

Students with Special Needs. While ST Math's self-paced curriculum can benefit all students, it might particularly benefit special education students, who often need to work at a slower pace or on a modified curriculum. The ST Math program permits classroom teachers to modify the grade-level material presented to students to match the requirements of their IEPs, allowing teachers to tailor instruction to special education students without pulling them from the classroom or into special learning groups. Gifted students, who often need more challenging material to prevent boredom and optimize achievement (McAllister & Plourde, 2008), can advance through the ST Math curriculum and explore optional challenge games. These forms of exceptionality are potential moderating variables to the main effect of ST Math. Therefore, we will investigate their potential aptitude-treatment interactions (ATIs), whereby students differentially benefit from the intervention.

Students with high or low initial spatial ability. An additional ATI might well be seen among students with differing levels of initial spatial ability. For the better part of a century, differential psychologists have known that people differ rather dramatically in their ability to generate and manipulate mental imagery. A spatial temporal approach to learning mathematics might interact with this individual difference variable. Here we must equivocate and approach the question of interaction with an open mind. It is not clear, *a priori*, whether spatial temporal mathematics instruction will fit best with students who are high in spatial temporal ability (because of a good "match" with cognitive characteristics) or whether it might especially help students who are low in spatial temporal ability (because explicit external representations might compensate for relative inability to generate those images internally) (Cronbach, 1975). Nevertheless, an ATI with spatial ability appears to be an obvious moderating variable, even if firm predications are elusive.

Methods

The Study Design

Strong research designs support valid inferences about the size of program effects, the warrants to make causal inferences, and the degree to which inferences drawn from limited samples justifiably generalize to larger populations. The current study design includes random assignment to treatment and control groups, the design component that goes furthest to supporting the causal inferences that lie at the heart of internal validity (Shadish, Cook, & Campbell, 2002). Additionally, a within school design was chosen to minimize correlated error terms between outcomes and the characteristics of students, teachers, schools, and neighborhoods.

Participants. The study population consists of two cohorts of ethnically diverse, majority Latino schools in Orange County, California. The demographics of the schools in Cohort 1, as measured in October 2008, were 2.3% black, .26% American Indian, 7.2% Asian, 1.7% Filipino, 81.8% Latino, .6% Pacific Islander, 5.6% White, with 84.4% on Free or Reduced Lunch, and 60.3% ELL (California Department of Education, 2008). Cohort 1 was selected to participate in the MIND Research Institute's Orange County Math Initiative (OC Math Initiative). This countywide initiative, supported by local business partners and the Orange County Department of Education, provides the ST Math instructional software without cost to low-performing schools. Over a three-year period, the ST Math program will be given to qualifying schools one or two grade levels at a time. To determine eligibility, every school in Orange County was ranked by its Academic Performance Index, which is based on a weighted composite of student scores on state-mandated standardized tests. Schools that fell into the lowest three deciles (155 elementary schools) were invited to participate in the OC Math Initiative. Among the qualifying schools, seventy-three schools applied to participate, and all 73 schools were accepted into the Initiative. A subset of 34 schools were eligible to participate in the study as they fully met the desired criteria: (1) They were not current users of ST Math, and (2) they signed a Letter of Intent to implement at grade levels (2/3 or 4/5) selected at random. Current ST Math users were excluded from the formal study because their prior exposure to the ST Math program would compromise the internal validity of the study. A few schools that had initially agreed to random assignment later reversed their position. Though they remain in the OC Math Initiative, they were excluded from this study.

Randomization. During the fall of 2008, schools in Cohort 1 were randomly assigned to one of two conditions: 18 Cohort 1 schools implemented ST Math at grades 2-3 and not in grades 4-5 (Group A), and another 16 schools implemented ST Math at grades 4-5 and not in grades 2-3 (Group B). Both sets of grade levels will be studied over a four-year period, with each school serving both as a treatment school at designated grades that use ST Math and as a control school for the grade levels in which the program is not implemented. Schools on the eligibility list were encouraged to apply for inclusion in Cohort 2 with the same restrictions and conditions. Sixteen Cohort 2 schools began implementation of the intervention at either grades 2-3 or 4-5 at the start of 2009-2010 school year.

As the schools progress through the study (Figure 5), they have the option of adding one or two grades of ST Math per year, such that schools who were initially in Group A (2nd and 3rd grades) during the 2008-2009 school year may implement the software in grade 4 the following year, while continuing to instruct 2nd and 3rd graders with ST Math. Those in Group B in 2008-2009 have the option of implementing ST Math for their kindergarteners in 2009-2010. All schools were initially given ST Math support and training for the first year of implementation, and can continue to receive support and training for the additional grades as scheduled by paying a $3500 renewal fee to MIND. Cohort 2 schools will follow the same design, but with the initial implementation year of 2009-2010. Figure 5, below, illustrates the path a student in Group A could take through conditions (ST or C for ST Math or Control, respectively) depending of student grade-level at the start of 2008-2009 (for Cohort 1) or 2009-2010 (for Cohort 2) school year. Group B paths would start with control conditions in grades 2 and 3.

Figure 5.
Group A’s Path From Year One to Exit from Study School

ST (Grade2) → ST (Grade 3) → ST (Grade 4) → ST (Grade 5)

ST (Grade 3) → ST (Grade 4) → ST (Grade 5) → Exit

C (Grade 4) → C (Grade 5) → Exit

C (Grade 5) → Exit

The Intervention

Students assigned to the intervention group spend two forty-five minute sessions each week working through the software under the direction of their classroom teacher. Although the students work at their own pace, teachers can help students make connections between ST Math games and the regular math curriculum taught in the classroom.. Teachers are available in the computer lab while the students work, and are able to view student data on progress through the games. Teachers are encouraged to review these student-level data and intervene if warranted by providing students with help to progress to the next level. Technical support and assistance from Client Support Specialists are available for teachers throughout the year.

Training is provided to teachers and school principals by MIND; optional training on implementing ST Math and integrating the program into the classroom curriculum was provided for teachers by staff from the Orange County Department of Education. Teachers were trained to recognize when students were "stuck" through both observation and the use of student data, and to respond by assisting individual students or targeting class lessons to support student progress through the software. Throughout the duration of the study, training will be offered and teacher participation will be tracked and analyzed for the possible impact of training variations on learning and motivation outcomes.

Assessment Measures

Several assessments will be used to measure students' annual mathematics and spatial reasoning skills.

Standardized Test Scores. In California, all students in grades 2-11 are tested on the California Standards Tests (CSTs) in the spring of each year. Tests assessing mastery of mathematics content standards are included, with each grade-level test assessing only standards in that grade level. Each test is developed by the Educational Testing Service (ETS) following a rigorous, multi-step validation process to confirm alignment to standards, depth and breadth of coverage, and cultural appropriateness. Tests are modified each year, with separate reliability and validity studies conducted annually. In 2007, the latest year for which this information is available, Cronbach's alphas in grade 2 and 3 mathematics were 0.93 and 0.94, respectively (Educational Testing Service, 2008). All students in the study take the CSTs each year; these scores will be utilized to assess student knowledge with respect to state content standards in mathematics.

Individualized Woodcock-Johnson Achievement and Cognitive Measures. Individual testing on math conceptual understanding, math problem solving, spatial aptitude, and mathematics-related motivation will be administered to randomly selected students in waves throughout the duration of the study. In early 2010, five students per treatment and control classroom were randomly selected for possible testing. Letters explaining the study and requesting parental permission to include their child in the evaluation were sent to the homes of each of the selected students. At the present time, individualized assessment is ongoing, and approximately 40 schools have been visited for testing. During school test visits, selected students with parent permission are randomly ordered for testing by trained faculty, graduate students, and undergraduate research assistants.

The individualized testing is conducted using the Woodcock-Johnson III (W-J III) battery. The W-J III has undergone multiple reliability and validity studies; stratified random sampling was utilized to obtain a national sample of 4,784 K-12 students in 2004. Each of the 42 subtests show reliabilities of 0.80 or higher; when combined into factors corresponding to the conceptual model for the test, most factors obtained reliability of 0.90 or above. Multiple validity studies have been conducted comparing scores on the W-J III to alternative measures of ability and achievement including the Wechsler Intelligence Scale for Children (WISC-III), the Naglieri Cognitive Assessment System, and the Kaufman Test of Educational Achievement. Each study resulted in appropriate convergent-discriminant validity on expected subtests.

For the purposes of this project, math concepts will be assessed from subtests that load on the W-J III's Gf factor (quantitative reasoning) and Gq ability: 10 Applied Problems, 18A Quantitative Concepts—Concepts, and 18B Quantitative Concepts—Number Series. Individual testing with these instruments allows a more targeted measure of the impact of ST Math on math concepts and reasoning than is provided by the CSTs. One assumption of the program developers is that ST Math supports increases in spatial reasoning. Student testing will therefore include the Block Rotation subtest of the W-J III.

Students' Mathematics Attitudes. In addition to cognitive measures, our individually-administered measures also include the assessment of student attitudes towards mathematics.

Attitudes were defined using the Eccles et al. expectancy-value model of motivation and are measured using a measure constructed from expectancy-value scales developed by Eccles (1993) et al. for the assessment of motivation. The validity and reliability of these scales has been well established in previous educational research, including within the context of mathematics learning (Wigfield & Eccles, 2000).

Within-Game Learning Patterns. A final set of student measures includes temporally-dense tracking of student activity. As each student engages the ST Math activities, real-time data on the choices (click-stream) and progress through the software are recorded and downloaded to MIND's servers. These data have been used previously to analyze student learning patterns within games and modules and to inform modifications to the game design (Hu et al., 2004, Martinez et al., 2008).

Teacher Efficacy. Teacher efficacy was assessed with the Mathematics Teacher Efficacy Belief Instrument (MTEBI, cite) modified to apply to current teachers. An initial version of this survey was administered in the fall of 2009 and is undergoing revisions for administration in the fall of 2010. Teacher demographic data (number of years' teaching experience, number of years in current assignment, undergraduate major) is also being collected for use in outcome analyses.

Data Analysis

Procedure for Initial Analysis

With one complete year of intervention for schools in Cohort 1, the analysis of California Standards Test (CST) data is expected to be informative. CST scores have very real consequences for schools and districts, and so information on student CST gains is extremely meaningful for school planning and decision-making. Evaluation of CST results also allows policymakers a view of ST Math's impact on a collection of skills selected as critical milestones by the State's standards committees. CSTs were also chosen for this initial analysis partly because of their availability for all students. CST results were aggregated by grade or subgroup within grade (gender, ELL status, free or reduced lunch status) for each study school. Scale scores were chosen as the unit of measurement for their increased precision and indication of more reliable treatment effects (National Center for Education Evaluation, 2009, p. 27-29), as well as enhanced comparability of scores across years. Future analysis using proficiency categories (advanced, basic, etc.) may prove valuable to provide a different measure of meaningful change; however, category cutoffs vary state-by-state, so proficiency measures may not prove to be generalizable (NCEE, 2009, p. 28).

Researcher files for study school Spring 2009 CST math subscores were downloaded from the California Department of Education website and checked for agreement with sample participant characteristics by researchers from MIND and UC Irvine. Available data include mean scale scores for subgroups of students within each school and grade that contained more than 10 students. The main analysis for all students without division by subgroup resulted in an N of 136 (4 grades at each of 34 schools). Not all subgroups were available at each school and grade level; the only subgroup with all schools reporting data was gender, where each school reported average scaled scores for boys and girls in each grade resulting in an N of 272 (34 schools, 4 grades, 2 group divisions per grade). Also evaluated was ELL status by looking at ELL students vs. non-ELL students (N=226) and economic status, as measured by comparing students on free or reduced lunch with those not on free or reduced lunch (N=210).

Results

2009 math scaled scores for Cohort 1 were regressed using OLS Regression onto intervention condition, controlling for the 2008 scaled scores of the same grade in the previous year, current grade level, and two school-wide demographic variables of interest, percent of ELL students and percent of students on free or reduced lunch. To account for between school differences, the model was clustered on school. Descriptive statistics for mean scores across grades and between treatment (ST Math) and control (no ST Math) are given in Table 1. Both treatment and control groups made year-to-year gains in CST math raw scores, but treatment group gains were consistently larger.

Table 1. Means of Mean Scaled CST Math Scores Across Grades by Treatment & Control

Grade	Grade 2		Grade 3		Grade 4		Grade 5	
Year	2008	2009	2008	2009	2008	2009	2008	2009
Treatment Mean Score (std dev)	345.06 (22.97) n=18	363.93 (14.9) n=18	343.28 (18.29) n=18	363.79 (21.97) n=18	351.31 (17.2) n=16	367.96 (15.44) n=16	344.87 (24.36) n=16	365.7 (30.12) n=16
	+18.87***		+20.51***		+16.65**		+20.83***	
Control Mean Score (std dev)	352.59 (23.88) n=16	363.3 (26.49) n=16	352.48 (19.13) n=16	369.28 (14.65) n=16	345.16 (14.25) n=18	356.76 (14.65) n=18	333.14 (23.95) n=18	347.37 (20.13) n=18
	+10.71*		+16.8***		+11.6**		+14.23*	

* $p < .05$ ** $p < .01$ *** $p < .001$ from paired t-test on difference of means

Regression results for Cohort 1 show an effect for ST Math that is significant at the .05 level with an average effect size of .37. While the previous year's test scores were significant predictors of 2009 math CSTs, neither current grade nor schoolwide demographic information were significant predictors of the scale scores. Regression coefficients and other relevant information are given in Table 2. ANOVAs were run to test possible interactions between the intervention and grade, gender, ELL status and economic status. Table 3 shows that no interactions were significant.

Table 2. Regression table of mean scaled score for 2009 CST regressed on ST Math condition, controlling for prior test scores at the grade level & demographics at school level

N=136 classrooms in 34 schools		**Coefficients**
ST Math	d B(se)	.37 5.98 (2.88)*
Grade	F	.67
Grade 3 v 2	β	.070
Grade 4 v 2	β	-.019
Grade 5 v 2	β	-.019
2008 Scaled Score	β	.641**
School-wide Controls		
% Eng Lang Learners	β	.020
% Free/RP Lunch	β	-.007
R-squared		.44
Adjusted R-squared		.41

*p ≤.05. **p ≤.001. All p values calculated considering cluster effect of school.

Table 3. F tests for interactions
none significant

Interaction	**F**	**p value**
GradexSTMath	.64	.59
ELLxSTMath	.01	.92
ELLxGradexSTMath	.36	.78
FreeLunchxSTMath	.69	.41
FreeLunchxSTMathxGrade	1.36	.26
GenderxSTMath	.09	.77
GenderxGradexSTMath	.78	.51

Discussion

Initial findings show promising results for ST Math in this randomized field trial. Aggregate student scores within each grade in each school show that ST Math positively impacts mathematics achievement as measured by the CSTs. The effect size of .37 reflects a non-trivial difference between ST Math and non-ST Math students and is considered an effect size between small and medium (Cohen, 1992). These data suggests that the spatial-temporal approach to mathematics instruction as expressed in the ST Math software can lead to gains in broad mathematics proficiency in the elementary school grades as measured by standards-based state assessments.

While interactions associated with the aggregated subgroup data were not significant, these results should not be taken as definitive even for this sample. The method of analysis for these data was less than ideal, as we were unable to include information from all schools for the ELL and Free/Reduced lunch subgroups. Also, aggregate data for these particular schools may be unlikely to detect interactions because of range restrictions, since the majority of Cohort 1 schools are over 50% ELL and have a student body over 80% of which participate in the free or reduced lunch program.

These initial aggregate findings may also be diluted by significant contamination across treatment and control grades. A certain amount of contamination was expected within the study design because of mixed-grade classrooms: a number of schools in Cohort 1 have classrooms made up of students in both grades 3 and 4. Students in these classrooms who are in designated 'control' grades in fact participate in ST Math lab time with their classmates. Additionally, Think Together, an after school care provider in one of the larger Cohort 1 districts, contracts separately with MIND for access to the ST Math software. Students within these schools are exposed to ST Math as part of their after school curriculum. Evidence for cross-grade contamination has been collected by the study team, and is being investigated for future analysis. Individual student test data and Fidelity of Implementation measures, both discussed in more detail below will also likely reduce the intrusion of these contaminations into future analysis of program effects. Even so, contamination across treatment and control conditions would reduce power overall, so the effect size of ST Math may well be conservatively estimated in this study.

Future Directions

Data collected to date have indicated that ST Math enhanced mathematics achievement test scores among second through fifth graders at 34 Orange County schools. Our next analysis step is to use individual student data to identify moderators and mediators of this effect. By shifting to the individual student as unit of analysis, it will be more clear who has received ST Math, to what degree they participated in the program, and if any implementation factors may have impacted results. Student background, including gender, ELL and economic status, as well as prior math achievement, motivational orientations and spatial aptitude will aid in identifying ATIs which may show that ST Math has stronger benefits for certain students. Data collection that will enable these analyses is already underway. By the close of the 2009-2010 school year, analysis of results from student Woodcock-Johnson and motivation measures, individual-level CST scores, and click-stream data from MIND will have begun.

In future analysis, longitudinal data will also be used to test whether multiple years of exposure to ST Math results in larger gains over time. Hierarchical Linear Modeling will shed light on how the intervention exerts effects over multiple levels of analysis—district, school, and grade. Furthermore, we intend to investigate the process-level mechanisms through which ST Math produces gains for learners by analytically unpacking the intervention to uncover potential mediating variables. Game-level data as well as aggregate qualitative data will help us determine how ST Math scaffolds learning of spatial techniques and mathematical concepts. For example, ST Math's structure of self-paced games within learning modules may increase student confidence for learning mathematics, leading to greater interest and future success in math learning. Whether and to what extent these motivational processes mediate the impact of ST math will be studied through measures of expectancy and value collected at multiple time points.

Finally, data on Fidelity of Implementation (FOI) will be helpful in determining how the intervention as intended actually becomes realized in the participating schools. Variations in FOI might well impact the effects across all measures of the study and constitute their own class of moderators. FOI measures will help to calibrate more exactly the variation in total time devoted to mathematics learning, a covariate that is not captured in the present analysis. To advance the project's research associated with FOI, MIND has worked with a team from The University of Chicago (Century, Freeman, Rudnick, & Leslie, 2007) to develop observation protocols and survey measures designed to determine how the critical components of ST Math are implemented across teachers and schools. We believe the rigorous analysis of FOI will prove valuable in determining which aspects of the program are the most necessary, and what implementation impediments are faced by ST Math schools.

Conclusion

US students are struggling to prepare for a world increasingly dependent on STEM proficiency in the workforce and the broader citizenry. The students in our study cohort of Orange County, California schools are no exception: Only 45 percent of students in this cohort scored at or above a proficient level on the 2008 math CSTs. Traditional math instruction, which typically relies heavily on verbal-analytic representations, is not producing the gains necessary for students to succeed as able math learners. ST Math provides a distinct approach that may meet the needs of larger numbers of students by tapping into their spatial ability and using that ability to build intuitive understandings of foundational math concepts such as proportionality and functions. Additionally, ST Math may reach students who the standard educational system has not served particularly well, including those for whom English is a second language. The approach may better address variation in spatial ability, including possible gender-related differences. Other pertinent dimensions of variation include socioeconomic status and speed of learning—some students may need self-pacing to either advance rapidly, or to revisit fundamentals. In addition to increasing math knowledge and reasoning, ST Math may provide students with academic successes that boost confidence, and an engaging math curriculum that spurs their desire to continue to learn.

References

AAUW (2010). *Why so Few? Women in Science, Technology, Engineering, and Mathematics.* Washington, DC: AAUW. Retrieved from AAUW website: http://www.aauw.org/research/whysofew.cfm

Association for Career and Technical Education (2007). *Issue Brief: Career and Technical Education's Role in Dropout Prevention and Recovery*. Available at: www.acteonline.org.

Barton, P. E. (2005). *One-third of a Nation: Rising Dropout Rates and Declining Opportunities.* Policy Information Center, Educational Testing Service, Princeton, NJ. Available at: www.ets.org/research/pic.

Brownell & Moser, (1949). *Meaningful versus mechanical learning: A study in Grade III subtraction*. Duke University Studies in Education, 8:1-207.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Century, J., Freeman, C., Rudnick, M., & Leslie, D. (2007). *A conceptual framework for fidelity of implementation of instructional materials*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY. March 28th 2008.

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3), 98-101.

Cronbach, L. J., (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*, 116-127.

Dennis, I., & Tapsfield, P. (1996) (Eds.). *Human abilities: Their nature and measurement*. Mahwah, NJ: Erlbaum.

Eccles-Parsons, J. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.). *Achievement and achievement motivations* (pp. 75-121). San Francisco, CA: W. H. Freeman & Co.

Educational Testing Service, (2008). *California Standards Tests (CSTs) Technical Report, Spring 2007 Administration*. Retrieved on May 21, 2008 from http://www.cde.ca.gov/ta/tg/sr/documents/csttechrpt07.pdf

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing Factorial Validity of the Mathematics Teaching Efficacy Beliefs Instrument. *School Science and Mathematics*, *100*(4), 194-202.

Fryer, R., & Levitt, S. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, *86*(2), 447-464.

Gandara, (2000). In the Aftermath of the Storm: English Learners in the Post-227 Era. *Bilingual Research Journal, 24*(1 & 2).

Garris, R., Ahlers, R., & Driskell, J. (2002). Games, motivation, and learning: A Research and practice model. *Simulation & Gaming*, *33*(4), 441-67.

Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005). *Reassessing U.S. International Mathematics Performance: New Findings from the 2003 TIMSS and PISA*. The American Institutes for Research. 1000 Thomas Jefferson Street, NW, Washington, DC 20007. Tel: 202-403-5000; Fax: 202-403-5001; e-mail: inquiry@air.org; Web site: http://www.air.org.

Grandin, T., Shaw, G., & Peterson, M. (1998). Spatial-Temporal versus Language-Analytic Reasoning: The Role of Music Training. *Arts Education Policy Review*, *99*(6), 11-14.

Gray, (1965). An experiment in the teaching of Introductory Multiplication. *The Arithmetic Teacher, 12*:199-52.

Graziano, Peterson & Shaw, (1999). Enhanced learning of proportional math through music training and spatial temporal training. *Neurological Research, 21*(2):139-52.

Hu, W., Bodner, M., Jones, E.G., Peterson, M.R., Shaw, G.L. (2004) Data mining of mathematical reasoning data relevant to large innate spatial-temporal reasoning abilities in children: Implications for data driven education. Paper presented at the Society for Neuroscience Abst 34th annual meeting.

Ingersoll, R. M. (2002). Out-of-Field Teaching, Educational Inequality, and the Organization of Schools: An Exploratory Analysis. University of Washington, Center for the Study of Teaching and Policy. Pp 17-18. Available at: http://depts.washington.edu/ctpmail.

Jepson & de Alth, (2005). English Learners in California Schools. Public Policy Institute of California.

Kilpatrick, Swafford, Findell, eds., (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kosslyn, S. (1980). *Image and mind*. Cambridge Mass.: Harvard University Press.

Linn, M.C., & Petersen, A.C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498.

Martinez, M. E., Peterson, M., Bodner, M. Coulson, A., Vuong, S., Hu, W., Earl, T., & Shaw G. L. (2008). Music training and mathematics achievement: A multiyear iterative project designed to enhance students' learning. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 396-409). New York: Routledge.

May, Henry, Irma Perez-Johnson, Joshua Haimson, Sattar, S., & Gleason, P. (2009). *Using State Tests in Education Experiments: A Discussion of the Issues* (NCEE 2009-013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.Available on the IES website at http://ncee.ed.gov.

McAllister, B., & Plourde, L. (2008). Enrichment Curriculum: Essential for Mathematically Gifted Students. *Education*, *129*(1), 40-49.

McCook, John, (2006). *The RTI Guide: Developing and Implementing a Model in Your Schools.* LRP Publications

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225.

National Academies of Science, (2007). Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. The National Academies Press, Washington, DC:1-13. www.nap.edu

National Association of Educational Progress, (2005). The Nation's Report Card, 2005. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006453

National Center for Education Statistics (2005). "International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results from the U.S. Perspective. Pp. 15 and 29. Available at: http://nces.ed.gov/pubs2005/2005003.pdf

National Center for Education Statistics (2009). *The Nation's Report Card: Mathematics 2009* (NCES 2010–451). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Mathematics Advisory Panel (2008). Foundations for Success: The Final Report of the National Mathematics Advisory Panel. http://www.ed.gov/about/bdscomm/list/mathpanel/index.html

National Research Council. Committee on Support for Thinking Spatially. (2006). *Learning to think spatially: GIS as a support system in the K-12 curriculum.* Washington, DC: National Academies Press.

No Child Left Behind Act of 2001, 20 U.S.C. § 1111 (2008).

Peterson M.R.[1], Balzarini, D., Bodner, M., Jones, E.G., Phillips, T., Richardson, D., & Shaw, G.L. (2004). Innate spatial-temporal reasoning and the identification of genius. *Neurological Research*, *26*, 2-8.

Programme for International Student Assessment (2007). Science Competencies for Tomorrow's World. Organisation for Economic Co-operation and Development. Available at www.pisa.oecd.org.

Shadish, W. R., Cook, D. T., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shaw, G., & Peterson, M. (2000). *Keeping Mozart in mind*. San Diego Calif.: Academic.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude of learning from instruction. *Journal of Educational Psychology, 78*(3), 347-376. “Achievement Gap” (citation)

Sorby, S. (2006). Developing 3D Spatial Skills for K-12 Students. *Engineering design graphics journal.*, *70*(3), 1-11.

Sorby, S. (2009). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education*, *31*(3), 459-480.

United States Department of Education (2010). *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act.* Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

What Works Clearinghouse, (2008). http://ies.ed.gov/ncee/wwc/reports/

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, 89(3), 451-469.

Wigfield, A., & Eccles, J. S. (2000). Expectancy‑value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68‑81.

Wright, W., & Li, X. (2008). High-Stakes Math Tests: How No Child Left Behind Leaves Newcomer English Language Learners behind. *Language Policy*, *7*(3), 237-266.