MainStreetAcademix
2 Rush Road - Henniker, New Hampshire
603-428-8706 - www.msanh.com

MathLine Program Evaluation:

A Summary of the Results of a Six-School, Randomized, Experimental Research Study of the Effects of MathLine on the Mathematics Achievement of Elementary School Students in Sullivan County, Tennessee.

2004-2005

Dr. William K. Preble
Alfred Newman
Dr. Zvi Szafran
Dr. Larry Taylor
9-1-05

Introduction

In 2004-2005, Main Street Academix conducted a rigorous program evaluation of Howbrite Solutions, Inc.'s MathLine multi-sensory math manipulative program. The study met the What Works Clearinghouse methodological 'gold standard', based on its randomized, pre-post, experimental design. Our research methodology and the results of the study are summarized here. This report is currently being revised for publication and is available upon request from the authors.

Methodology

The MathLine evaluation was designed to determine whether or not the use of the MathLine number line tool would improve student learning in mathematics. The study was conducted in six randomly selected schools in East Tennessee. Three elementary schools were randomly assigned to an experimental condition and three served as the control group.

<u>Research Questions</u>

1. To what extent are there differences in mathematical skills and knowledge, as measured by the AGS Standardized Mathematics Test, between a randomly selected treatment or experimental group (EXP)—students who received mathematics instruction using the MathLine multi-sensory math manipulative program--- and a control group (CON) of students who do not use MathLine?
2. What are the effects of the use of the MathLine multi-sensory math manipulative program on student learning in mathematics, as measured by the AGS Standardized Mathematics Test, in three randomly selected elementary schools (grades 1-4).
3. To what extent do teachers who have received professional development training use MathLine "with fidelity" or "as directed" to successfully supplement their mathematics instruction?
4. How do teachers and students who have the opportunity to use MathLine describe their experiences?

<u>Selection and Assignment of Experimental Units</u>

Eighteen elementary schools in Sullivan County, Tennessee were paired on the basis of several demographic factors including the size of the school (population), the percentage of students receiving free and reduced lunch, and the school's performance on the state mathematics test. These pairs were then blocked by state mathematics assessment performance into three levels, i.e. high performing schools, middle performing schools and low performing schools with three pairs of schools in each block. One pair from each block was selected at random. One school from each selected pair was, then, randomly assigned to either an experimental group (EXP) or control group (CON).

<u>Training of Teachers</u>

MathLine materials (the tool, teacher manuals, student workbooks) were initially distributed and a standard, half-day professional development workshop on teaching with MathLine was conducted by a Howbrite Solutions trainer in August, 2004 for all grade 1-4 teachers in the three experimental schools. Teachers received no more formal training after the half-day workshop.

Pre-test and Post-test Administration

Students in grades 1-4 were given a grade-level baseline test of mathematical skills, the G*MADE Tests by AGS Publishing, in October, 2004 (PreTest) and a parallel test (PostTest) in May, 2005. These valid, reliable tests are norm-referenced, NCTM and state standards-based, and compliant with the No Child Left Behind Act of 2001.

The G*MADE test produces a set of raw scores for total mathematics achievement (TOTAL). The standardized test contains three subtests consisting of items designed to measure specific skills: Concepts and Communication (Concepts), Operations and Computation (Operations), and Process and Applications (Process). EXP and CON were compared for mean improvement of score (Δ Score) in the TOTAL and three subtests.

The G*MADE pre-test of student mathematics achievement was administered to 1,145 grade 1-4 students in all six schools (three experimental and three control). 1145 of 1274 students (90%) who took the PreTest also completed the PostTest. No significant disruption or attrition of subjects occurred during the study.

Fidelity of Implementation

'Fidelity of Implementation' (FOI) observations of teachers' use of the MathLine number tool in mathematics instruction were conducted in the experimental schools in October and November, 2004, and January, 2005. To maintain comparability, basic observations of teachers (while they were teaching mathematics) and limited interviews took place as well in the control schools. These FOI observations in the experimental schools were done to assess the extent to which the MathLine tools were being used as directed. Teachers and students participating in the study were interviewed in January and April, 2005 to obtain teacher and student narratives about the effects of using MathLine in the classroom.

A classroom observation tool was designed by MSA to assess the ways in which classroom teachers use MathLine in their classrooms. A second rubric was designed to assess the overall classroom environment observed in each classroom. Classroom observers were trained in the use of the tools and these tools were pilot tested in several classrooms whereupon a test of inter-rater reliability was completed.

Section A of the Fidelity of Implementation Instrument (Exhibit 1) consisted of seven questions designed to measure how well MathLine was used in the classroom. This section was only used in observations at experimental schools. If the MathLine tool was effective, classrooms with strong implementation scores would be expected to show greater differences in pre-test/post-test scores than classrooms with weak implementation scores.

Section B of the instrument consisted of six questions designed to measure the overall classroom environment. This section was used at both the experimental and control schools. It was expected that data from Section B could be analyzed to separate differences in learning due to using MathLine (factor a) from differences in learning due to superior (or inferior) classroom environment (factor c).

Inter-rater Reliability of the Fidelity of Implementation Instrument

The inter-rater reliability of the Fidelity of Implementation instrument was measured during the first set of classroom observations. A total of 24 classrooms at all grade levels participating in this study were simultaneously but independently observed by two representatives of Main Street Academix. Inter-judge reliability scores were initially found to be high (85%+), and further minor refinements in the FOI tool increased the reliability scores.

Qualitative Results

We agree with Michael Quinn Patton (2002) that effective program evaluations 'use both qualitative and quantitative methods to understand the nuances and idiosyncrasies of particular evaluation questions (p.68)'. Both qualitative and quantitative data were collected during this study. The use of a multiple methods design provided a variety of sources of evidence about the effects of MathLine on students and teachers. As we compared the findings from these mixed methods we were encouraged to see that the results all pointed in the same direction.

Qualitative data collected from interviews with teachers were generally very positive and pointed to specific areas where teachers felt that MathLine was making a difference with their students: The quotes below are representative of some of the positive teacher comments that were made about their use of MathLine:

- *"Several students in my class struggle with math in general. Having a tool such as MathLine shows them math with their eyes." (3rd grade teacher)*

- *"It is very good for students with problems and it helps students to understand the concept. (3rd grade teacher)*

- *" It is utilized best to introduce new skills that are unfamiliar to the students." (4th grade teacher)*

- *"It did help students see 'rounding' better…especially at first when they could see the blue and red rings. It didn't take long." (4th grade teacher)*

- *"For those who had a mental block memorizing facts, it has helped them see the connection."(4th grade teacher)*

From these interviews (conducted during the observations measuring fidelity of implementation), we concluded that most teachers felt that their use of MathLine went well. Several teachers in the lower grades reported that they didn't use MathLine all the time (and felt it worked best as a supplement), because they were afraid it would become a "crutch" for the students. Various teachers pointed out its advantages in teaching odds and evens, skip counting, basic addition and subtraction and rounding. Teachers in the 4th grade (and to some extent in the 3rd) felt that their students had already "outgrown" the use of MathLine, and only used it for remedial students. Many teachers did not seem to understand the normal student learning curve for manipulatives, i.e., students will use them intensively in the beginning, and then set them aside when they are comfortable with the concept.

Teachers reported many positive experiences using MathLine. Most related stories about individual students who were unable to grasp a concept (such as subtraction, counting by 5's, rounding, etc.) until MathLine made it "concrete" to them. This was especially true for weaker and remedial students.

Negative experiences mainly centered around the "becoming a crutch" issue. Teachers in grade 2 felt that the 100-counter MathLine instrument would be more appropriate for their use than the one supplied. Several commented on the lack of sturdiness of the instrument. Suggestions for improvement included more teacher training, more advanced applications, using the 100-counter MathLine in the 2nd grade, and having better suction cups on the instrument.

We also interviewed students about what they thought about using their MathLines. Below are a number of representative student comments:

Some first graders told us:

- *"It helps me with counting money."*
- *"It helps with hard problems."*
- *"It's exciting because you can learn."*

Second and third graders said:

- *"It helps us learn."*
- *"It helps me to add and subtract."*
- *"It makes it easier to do Math."*
- *"The colors on the rings make it easier to count."*

One rather articulate 4th grader told us:

- *"It sparks my interest and makes me more interested in math."*

Several students also told us that they didn't use MathLine because they didn't need it.

These qualitative results describe how teachers and students felt about their use of MathLine. While these descriptive results were generally very positive, such qualitative data do not 'prove' that using MathLine helps students to learn math concepts, skills, and operations more successfully than students who do not have the use of these mathematics tools. These quotes do, however, reveal the mostly positive feelings and thinking of students and teachers about the use of MathLine. These interview data are, in fact, consistent with the results gleaned from more rigorous statistical research methods reported below.

Quantitative Results

First Observation - Fidelity of Implementation Findings
Analysis of the data obtained from the first observations revealed that all experimental schools were relatively similar in their use of the MathLine manipulative, and that across these schools the classroom environments were also similar. Within schools, some teachers were observed to use the MathLine manipulative more creatively than others, but in general teachers were more similar than different in their use of MathLine.

Second Observation – Fidelity of Implementation Findings
The average scores for the second observation were generally lower in every case than the first observation. This would indicate that some teachers were using MathLine less and with fewer students on a regular basis, as their students had already mastered the mathematical concepts. The average FOI score increased for the first grade, but this was a result of the lowest score moving up as teachers gained familiarity with the tool. The results for the 1st through 3rd grades were quite similar, with the 4th grade results being significantly lower. This finding corroborated the comments that we received from teachers that the instrument would seem to be more useful in grades K-3 than in grade 4.

Relatively small differences between schools were observed. This indicates that the experimental schools had implemented MathLine in a similar way, and were generally familiar with its use.

Third Observation – Fidelity of Implementation Findings
The average scores for the third observation (March 2005) were generally higher than the second observation (January 2005), and comparable to the first observation (October 2004). This would indicate that teachers used MathLine on a regular basis in the beginning, a lull then occurred, and then used MathLine more regularly once again, perhaps in anticipation of the upcoming TCAP exams. The 1st grade scores improved significantly, indicating greater teacher familiarity with MathLine.

Relatively small differences between schools were observed in the third observation scores, indicating that the schools had implemented MathLine in a similar way, and were generally familiar with its use.

Descriptive Results on the G* MADE Standardized Mathematics Test (AGS)

The *percentage differences* between the pre-test and post-test were calculated for each school at each grade level. Percentage differences for Sub-Tests were calculated in the areas of “Concepts”, “Operations” and “Process”, using the same classification scheme as defined by the AGS G*MADE Tests. The results are shown in Table I, below.

Overall Results
The greatest percentage improvement on the G*MADE was shown in the 1st grade, for both experimental and control schools. The 2nd and 3rd grade results also showed

marked improvements in both the experimental and control schools, but these gains were smaller than the 1st grade gains. The 4th grade improvements were smallest for both experimental and control schools.

At each grade level, the strongest percentage improvements came in the sub-area of "Operations", followed by "Process". Conceptual improvements (Concepts) were smallest, but that may have been because the pre-test results in the conceptual sub-area were highest. This indicates that on the pre-test, while students understood the basic concepts of addition, subtraction, multiplication and division, they were not effective problem-solvers using these skills. MathLine's greatest positive effect was in improving problem solving skills. Conceptual understanding, which was already good, therefore improved at a smaller rate.

Grade-Level Results

At the 1st grade level, the experimental schools showed markedly stronger percentage improvements in every sub-area than the control schools, with impact factors (defined as the percentage improvement of the experimental schools divided by the percentage improvement of the control schools) ranging from 1.43 to 1.62, with an average of 1.50. This indicates that the students using MathLine improved their scores an average of 50% more than students not using MathLine. The greatest impact factors were in the area of "Concepts" and "Operations".

At the 2nd grade level, the experimental schools again showed markedly stronger percentage improvements in every sub-area, with impact factors ranging from 1.58 to 1.70, with an average impact factor of 1.59. This indicates that the students using MathLine improved their scores an average of nearly 60% more than students not using MathLine. The greatest impact factor was in the area of "Concepts".

At the 3rd grade level, the experimental schools again showed markedly stronger percentage improvements in every sub-area, with the largest impact factors seen in this study, ranging from 1.50 to 2.05, with an average impact factor of 1.73. This indicates that the students using MathLine improved their scores an average of nearly 75% more than students not using MathLine. The greatest impact factor was in the area of "Operations".

At the 4th grade level, the experimental schools and control schools showed no significant differences in percentage improvements, and the percentage improvements were the smallest for any grade level. This was expected, as the Fidelity of Implementation instruments showed the greatest degree of usage of MathLine in Grades 1-3, with many of the 4th grade teachers generally teaching math skills without the routine use of Math Line , in many cases allowing it only for the weakest students. The impact factors ranged from 0.90 to 1.17, with an average of 0.99. This indicates that the students using MathLine improved their scores by the same factor as students not using MathLine (the difference was 1% favoring the control schools).

Chart I

Percentage Improvement Between Pre- and Post-Test

	Concepts	Operations	Process	TOTAL
Blountville Gr. 1 EXP	30.41	158.46	48.36	58.88
Valley Pike Gr. 1 EXP	32.74	257.01	65.31	78.52
Sullivan, Gr. 1 EXP	40.66	219.76	63.73	75.23
Average, Gr. 1 EXP	**34.60**	**211.74**	**59.13**	**70.88**
Cedar Grove Gr. 1 CON	23.84	162.67	52.81	55.40
Miller Perry Gr. 1 CON	17.85	127.81	31.74	41.77
Mary Hughes Gr.1 CON	22.37	104.61	39.11	44.57
Average, Gr. 1 CON	**21.35**	**131.70**	**41.22**	**47.25**
Impact Factor, Grade 1	***1.62***	***1.61***	***1.43***	***1.50***
Blountville Gr. 2 EXP	21.28	57.54	51.19	39.22
Valley Pike Gr. 2 EXP	35.54	93.26	89.95	66.97
Sullivan, Gr. 2 EXP	9.63	75.92	53.01	37.87
Average, Gr. 2 EXP	**22.15**	**75.57**	**64.72**	**48.02**
Cedar Grove Gr. 2 CON	13.14	30.83	18.79	19.71
Miller Perry Gr. 2 CON	20.11	61.14	50.29	41.50
Mary Hughes Gr.2 CON	5.81	51.87	49.07	29.24
Average, Gr. 2 CON	**13.02**	**47.95**	**39.38**	**30.15**
Impact Factor, Grade 2	***1.70***	***1.58***	***1.64***	***1.59***
Blountville Gr. 3 EXP	16.64	193.69	89.18	68.71
Valley Pike Gr. 3 EXP	22.83	98.97	53.72	50.50
Sullivan, Gr. 3 EXP	12.37	129.47	51.18	47.71
Average, Gr. 3 EXP	**17.28**	**140.71**	**64.69**	**55.64**
Cedar Grove Gr. 3 CON	13.27	55.82	41.59	30.96
Miller Perry Gr.3 CON	11.23	70.25	38.57	33.08
Mary Hughes Gr.3 CON	10.13	80.25	32.61	32.47
Average, Gr. 3 CON	**11.54**	**68.77**	**37.59**	**32.17**
Impact Factor, Grade 3	***1.50***	***2.05***	***1.72***	***1.73***
Blountville Gr. 4 EXP	8.45	92.30	53.11	38.67
Valley Pike Gr. 4 EXP	6.44	37.82	18.57	18.16
Sullivan, Gr. 4 EXP	1.23	41.50	29.68	18.75
Average, Gr. 4 Exptl	**5.37**	**57.21**	**33.79**	**25.19**
Cedar Grove Gr. 4 CON	0.68	28.04	23.73	13.34
Miller Perry Gr. 4 CON	6.60	50.64	53.27	31.66
Mary Hughes Gr.4 CON	8.48	68.20	35.35	31.68
Average, Gr. 4 CON	**5.25**	**48.96**	**37.45**	**25.56**
Impact Factor, Grade 4	***1.03***	***1.17***	***0.90***	***0.99***

Summary of Descriptive Results

These quantitative data indicate that the greatest improvement effect (impact factor) of using MathLine is in grades 1-3, with grade 4 showing essentially no effect. This correlates well with teacher comments indicating that MathLine should be implemented in grades K-1 (for the 25 bead instrument) and grades 2-3 (for the 100 bead instrument), rather than in grades 1-4. There is, of course, strong circularity to these results: 4th grade teachers did not heavily use MathLine since they felt it was inappropriate at that grade level. Therefore, the small amounts they used MathLine resulted in little quantitative impact.

While the 4th grade results do not indicate any effect for MathLine use, they do illustrate the reliability of the overall experimental design. The experiment was designed, by matching schools of similar size and demographic nature and by measuring differences in teaching styles within classrooms, to equalize all causes of improvement other than the use of the MathLine instrument. In the 4th grade, where the MathLine instrument was hardly used, the overall impact factor of 0.99 indicates the extremely strong comparability of the experimental and control schools and the teaching methods used (1.00 would indicate perfect comparability).

We therefore conclude that the experimental design was substantially accurate, and that when used, MathLine has a strong affect (50% or more) on the improvement students show in mathematics in all sub-areas at the 1st-3rd grade levels.

Tests of Statistical Significance

A third method of evaluating the effects of MathLine was to conduct more rigorous statistical analyses of students' test score data. We conducted t-tests to evaluate the statistical differences between experimental and control groups.

The distributions for all pre-test scores and mean score changes on posttests were examined and found to be essentially normal. Weighted *t*-tests were performed at the school level using the school means for PreTest and Δ Score (Delta Score means the Pre-Posttest Change Score) as independent samples. Mean scores for Experimental and Control groups were compared for mean improvement in the Mathematics TOTAL SCORE and within each of the three subtests.

On the Pre-Test, the experimental schools exhibited ***lower*** mean scores in all subtests. Nonetheless, these lower mean scores of the experimental schools were not significantly different from the control schools' mean scores. This difference in pre-test scores on the G* MADE was somewhat surprising given the careful nature of paired, randomized sampling of experimental and control schools who were matched due to previous similarities of standardized test scores in mathematics. (see details of Pre-Test scores on Table 1 below).

Weighted t-Tests for PRETEST

EXP School	Concepts	Operations	Process	TOTAL
BLV	20.21	9.08	11.73	41.03
SUL	20.81	9.94	12.67	43.36
V P	19.93	9.97	12.87	42.77
Mean	**20.42**	**9.53**	**12.24**	**42.17**
s^2 =	**21.63**	**154.95**	**72.13**	**654.97**
s =	**4.65**	**12.45**	**8.49**	**25.59**

CON School	Concepts	Operations	Process	TOTAL
C G	20.59	10.45	12.74	43.78
M H	22.02	11.40	14.30	47.71
M P	21.89	11.67	14.56	47.98
Mean	**21.71**	**11.39**	**14.18**	**47.20**
s^2 =	**12.41**	**93.42**	**52.28**	**417.60**
s =	**3.52**	**9.67**	**7.23**	**20.44**

	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
s_{pooled} =	**4.13**	**11.14**	**7.89**	**23.16**
t =	-0.38	-0.20	-0.30	-0.27
abs(t) =	**0.38**	**0.20**	**0.30**	**0.27**
p =	**0.722**	**0.848**	**0.778**	**0.803**

Table 1

Table 2 below shows that the experimental schools demonstrated significant

Weighted t-Tests for Δ Score

EXP School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
BLV	3.60	10.25	6.86	20.71
SUL	3.33	10.04	7.13	20.51
V P	5.17	11.30	8.17	24.63
Mean	**3.68**	**10.29**	**7.12**	**21.09**
s^2 =	**1.17**	**0.54**	**0.58**	**6.53**
s =	**1.08**	**0.74**	**0.76**	**2.56**

CON School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
C G	20.59	10.45	12.74	43.78
M H	22.02	11.40	14.30	47.71
M P	21.89	11.67	14.56	47.98
Mean	**21.71**	**11.39**	**14.18**	**47.20**
s^2 =	**0.68**	**0.48**	**1.12**	**6.28**
s =	**0.83**	**0.69**	**1.06**	**2.51**

	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
s_{pooled} =	**0.77**	**1.06**	**0.80**	**2.37**
t =	1.71	3.05	2.45	2.69
abs(t) =	**1.71**	**3.05**	**2.45**	**2.69**
p =	**0.081**	**0.019**	**0.035**	**0.027**

Table 2

improvement in scores over their control counterparts. Weighted *t*-tests (one-tailed) for Operations (p = 0.019), Process (p=0.035) and TOTAL (p=0.027) were all statistically significant at the $\alpha = 0.05$ level. Weighted *t*-test (one-tailed) for Concepts was not statistically significant (p=0.08), but EXP scores still showed greater improvement. When we analyzed the differences in the changes in students' scores by Grade, we found that the significant values occurred mainly in 1st and 3rd grades and in the ΔConcepts in 1st Grade. (see Appendix.) These strongly significant *p*-values were major contributors to the statistical significance seen in experimental vs. control overall. Looking at the FOI scores by Grade, we noted that the 1st and 3rd grade mean FOI scores were also higher than 2nd and 4th grade mean FOI scores. These FOI differences were, however, not statistically significant.

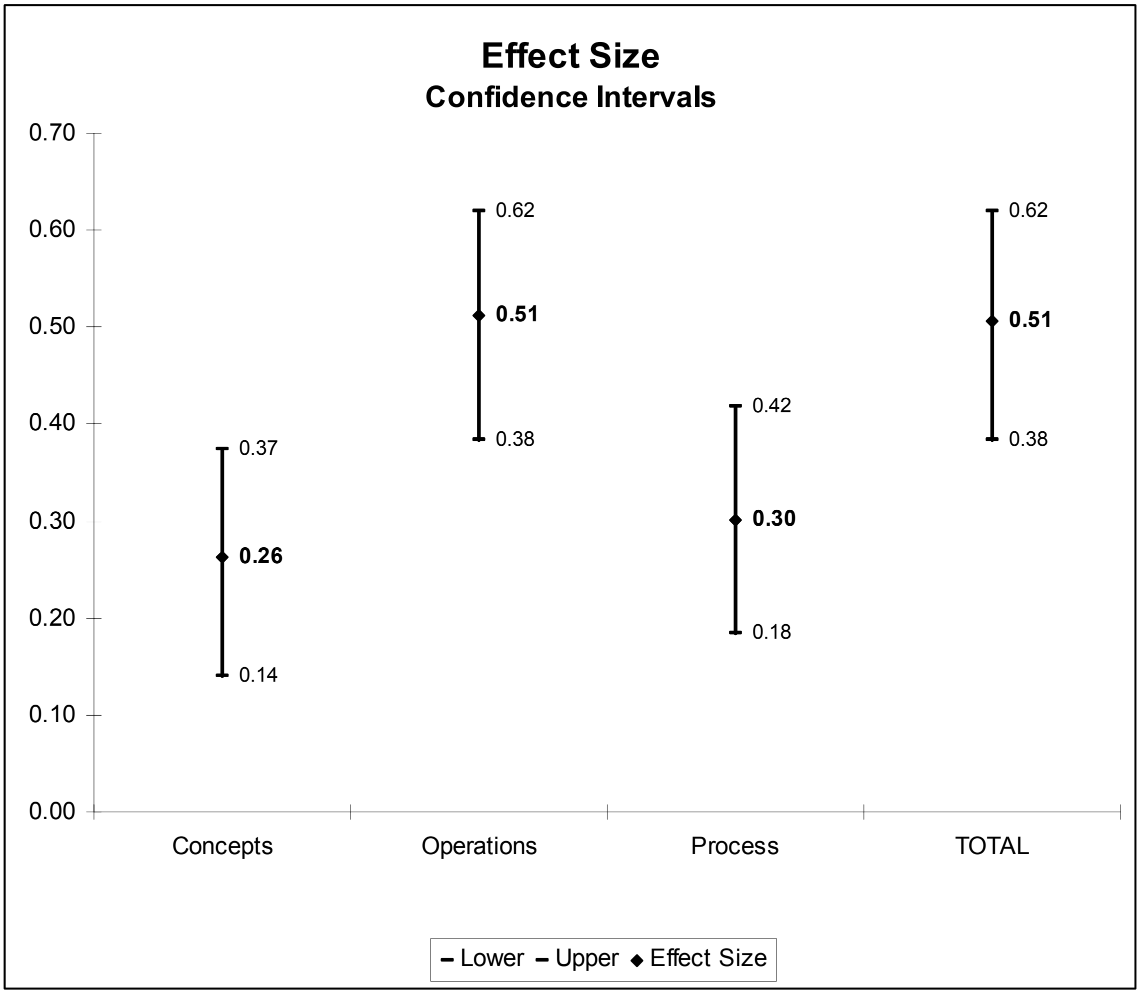
Grade	Mean FOI
1st	17.3
2nd	15.7
3rd	17.0
4th	15.1

Effect Size Analysis

The What Works Clearing House has determined that one of the most important measures for determining differences between experimental and control groups in program evaluation studies, such as this one, is a measure called, “effect size” or ES. Cohen (1988) hesitantly defined effect sizes as "small, *d* = .2," "medium, *d* = .5," and "large, *d* = .8".Effect sizes can be interpreted in terms of the percent of non-overlap of the treated group's scores with those of the untreated group. An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group, there is 0% of non-overlap. An ES of 0.8 indicates a non-overlap of 47.4% in the two. When the treatment group’s pre-post-test scores are significantly different (non-overlap) than the control group, it seems reasonable to assume that there is a difference between the two groups.

Below is a summary of effect size analysis of the experimental and control group in our study.

DATA ENTRY							STANDARDISED EFFECT SIZE					
Outcome measure	Treatment group			Control group			Effect Size	**Bias corrected (Hedges)**	Standard Error of E.S. estimate	Confidence Interval for Effect Size		Effect Size based on control group SD
	mean	n	SD	mean	n	SD				lower	upper	
Δ Concept	3.68	516	4.34	2.6	628	4.11	0.26	**0.26**	0.06	0.14	0.37	0.26
Δ Operation	10.29	516	5.38	7.65	628	5.16	0.50	**0.50**	0.06	0.38	0.62	0.51
Δ Process	7.12	516	5.25	5.53	628	5.28	0.30	**0.30**	0.06	0.18	0.42	0.30
Δ TOTAL	21.09	516	10.54	15.86	628	10.34	0.50	**0.50**	0.06	0.38	0.62	0.51



It would appear from these analyses, that, using Cohen's definition of the relative strength of an effect size score, there is a 'medium' difference (d=.51) in "Concepts" and "TOTAL" scores on the G*MADE. This would support the previous conclusions that the use of MathLine did make a positive contribution to student gains in academic achievement when compared to the control group.

Conclusions

The results from teacher and students interviews, the descriptive data showing percentage differences between the pre-test and post scores of experimental and control groups, the Weighted *t*-tests performed at the school level using the school means to examine Pre-Posttest Change Scores, and the Effect Size analyses all yielded similarly positive results.

They show that students in experimental schools started the study in August, 2004, by demonstrating lower Pre-Test scores than students in the control group. By the end of April, 2005, the students in the experimental group had made significantly greater overall academic gains than control group students.

While Post-Test score differences for both groups were not significantly different in April, 2005 the results indicated that the experimental group managed to 'catch up' academically with the control group over the course of the year using MathLine. These results indicate that students who used MathLine made significantly greater academic progress over the course of an academic year than did students who did not have MathLine as a learning tool in their classrooms; essentially bridging the achievement gap that was evident in the pre-test.

It is unclear why schools initially selected and matched based upon their previous math scores on state tests would have differing pre-test results. This leads to two possible conclusions about our statistically significant findings.

- MathLine significantly improves the performance of elementary school students, or
- It was "easier" for students in the experimental schools to show marked gains, since they started from a somewhat lower score. The larger gain in the experimental schools may therefore be an artifact of this initial difference.

Students of teachers who used MathLine "as directed" or those teachers who demonstrated the greatest levels of 'fidelity of implementation', based on classroom observations, had higher mathematics scores than students whose teachers used MathLine with lower levels of fidelity, although these positive differences were not statistically significant in this study.

When taken all together, the evidence generated by this research seems clear and compelling. It is our judgment that the first conclusion is accurate, that MathLine did significantly improve test scores in mathematics (understanding of basic mathematical constructs) of elementary school students in Tennessee.

Thanks to all those teachers, principals, and students from both the experimental and control schools in Sullivan County, Tennessee who participated in this study and to Carole Briggs, Jack Barnes, and Janie Barnes for their support of this research.

Citations:

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Patton, M. Q. (2002).Qualitative Research and Evaluation Methods. London, England: Sage Publications.

Appendix

IMPROVEMENT - Grade 1

	EXP					CON			
School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL		ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
Blountville	4.81	10.33	6.07	21.21	Cedar Grove	4.43	9.65	7.22	21.30
Sullivan	7.08	12.06	8.15	27.29	Miller Perry	3.66	10.02	5.43	19.10
Valley Pike	6.53	16.18	10.41	33.12	Mary Hughes	4.40	9.26	6.28	19.93
Total	**5.94**	**11.83**	**7.49**	**25.26**	Total	**3.98**	**9.75**	**5.93**	**19.67**
	1.29	2.23	2.02	16.42		0.21	0.01	1.65	2.68
	1.30	0.06	0.43	4.12		0.11	0.07	0.26	0.32
	0.34	18.91	8.53	61.71		0.17	0.25	0.12	0.07
s^2 =	**1.47**	**10.60**	**5.49**	**41.12**	s^2 =	**0.24**	**0.16**	**1.01**	**1.53**
s =	**1.21**	**3.26**	**2.34**	**6.41**	s =	**0.49**	**0.41**	**1.01**	**1.24**
s_{pooled} =	**0.92**	**2.32**	**1.80**	**4.62**					
	0.75	1.89	1.47	3.77					
t =	2.60	1.10	1.06	1.48					
abs(t) =	**2.60**	**1.10**	**1.06**	**1.48**					
p =	**0.030**	**0.167**	**0.175**	**0.106**					

IMPROVEMENT - Grade 2

	EXP					CON			
School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL		ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
Blountville	4.55	6.97	7.00	18.52	Cedar Grove	3.74	4.87	4.04	12.65
Sullivan	2.19	8.28	8.13	18.59	Miller Perry	3.62	7.81	7.11	19.11
Valley Pike	6.38	10.38	10.63	27.38	Mary Hughes	1.28	6.13	6.74	14.15
Total	**3.78**	**7.95**	**7.93**	**19.66**	Total	**2.96**	**6.90**	**6.56**	**16.74**
	0.59	0.96	0.86	1.30		0.60	4.12	6.33	16.73
	2.55	0.11	0.04	1.13		0.44	0.83	0.30	5.61
	6.73	5.90	7.26	59.58		2.82	0.59	0.03	6.71
s^2 =	**4.93**	**3.49**	**4.08**	**31.00**	s^2 =	**1.93**	**2.77**	**3.33**	**14.52**
s =	**2.22**	**1.87**	**2.02**	**5.57**	s =	**1.39**	**1.66**	**1.83**	**3.81**
s_{pooled} =	**1.85**	**1.77**	**1.93**	**4.77**					
	1.51	1.44	1.57	3.90					
t =	0.54	0.72	0.87	0.75					
abs(t) =	**0.54**	**0.72**	**0.87**	**0.75**					
p =	**0.308**	**0.255**	**0.216**	**0.248**					

IMPROVEMENT - Grade 3

	EXP					CON			
School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL		ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
Blountville	3.59	14.46	9.60	27.65	Cedar Grove	2.74	6.59	4.52	13.85
Sullivan	3.28	11.91	6.70	21.89	Miller Perry	2.29	7.92	5.27	15.48
Valley Pike	4.94	11.29	7.65	23.88	Mary Hughes	2.35	8.93	5.03	16.30
Total	**3.65**	**13.14**	**8.32**	**25.11**	Total	**2.38**	**7.95**	**5.08**	**15.41**
	0.00	1.72	1.65	6.42		0.13	1.84	0.31	2.43
	0.14	1.51	2.61	10.37		0.01	0.00	0.04	0.01
	1.66	3.42	0.45	1.52		0.00	0.95	0.00	0.79
s^2 =	**0.90**	**3.33**	**2.36**	**9.15**	s^2 =	**0.07**	**1.40**	**0.18**	**1.61**
s =	**0.95**	**1.82**	**1.54**	**3.03**	s =	**0.26**	**1.18**	**0.42**	**1.27**
s_{pooled} =	**0.70**	**1.54**	**1.13**	**2.32**					
	0.57	1.25	0.92	1.89					
t =	2.23	4.14	3.53	5.12					
abs(t) =	**2.23**	**4.14**	**3.53**	**5.12**					
p =	**0.045**	**0.007**	**0.012**	**0.003**					

IMPROVEMENT - Grade 4

	EXP					CON			
School	ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL		ΔConcepts	ΔOperations	ΔProcess	ΔTOTAL
Blountville	1.72	8.69	4.57	14.98	Cedar Grove	0.39	3.45	2.58	6.42
Sullivan	1.44	8.53	5.69	15.66	Miller Perry	1.16	6.13	5.49	12.78
Valley Pike	1.30	4.50	1.30	7.10	Mary Hughes	1.70	8.43	4.08	14.22
Total	**1.57**	**8.31**	**4.82**	**14.69**	Total	**1.13**	**6.11**	**4.56**	**11.80**
	0.02	0.15	0.06	0.08		0.53	7.06	3.95	28.90
	0.02	0.05	0.76	0.94		0.00	0.00	0.86	0.96
	0.07	14.49	12.40	57.67		0.33	5.38	0.23	5.84
s^2 =	**0.06**	**7.34**	**6.61**	**29.34**	s^2 =	**0.43**	**6.22**	**2.52**	**17.85**
s =	**0.24**	**2.71**	**2.57**	**5.42**	s =	**0.66**	**2.49**	**1.59**	**4.22**
s_{pooled} =	**0.50**	**2.60**	**2.14**	**4.86**					
	0.40	2.13	1.74	3.97					
t =	1.09	1.03	0.15	0.73					
abs(t) =	**1.09**	**1.03**	**0.15**	**0.73**					
p =	**0.168**	**0.180**	**0.445**	**0.253**					