An Evaluation of the 2005-2006 Reasoning Mind Project

Submitted to

Reasoning Mind, Inc.
410 Pierce Street, Suite 208
Houston, Texas 77002

Submitted by

W. A. Weber
College of Education
University of Houston

December 2006

AN EVALUATION OF THE 2005-2006 REASONING MIND PROJECT

Introduction

This report provides a description of the results obtained from analyses of the data collected for an external evaluation of the Reasoning Mind Project—a project intended to enhance student achievement in mathematics which was field tested in ten schools in the Houston metropolitan area during the 2005-2006 school year. First, because the very positive results obtained from an evaluation of the Reasoning Mind Pilot Project which was conducted during the Spring 2003 semester set the stage for the implementation of the Reasoning Mind Project during the 2005-2006 school year, this report begins with a very brief discussion of the results obtained from that evaluation (Weber, 2003). Second, the report provides a brief description of the evaluation plan used to evaluate the Reasoning Mind Project during the 2005-2006 school year. Third, the report presents a summary of the results obtained from analyses that were designed to describe the extent to which the Reasoning Mind Project was effective in enhancing student achievement in mathematics. Fourth, the report presents a summary of the results obtained from subsequent analyses—post hoc analyses—that were undertaken in an effort to provide additional explanations regarding the results obtained from the analyses that examined the effect of the Reasoning Mind Project. Fifth, the report provides a summary of the results obtained from analyses of student attitudes toward the Project. Sixth, the report summarizes the results. Seventh, the evaluator offers some observations. Eighth, the report discusses the perceptions of Reasoning Mind Project staff regarding the most serious problems that affected the outcomes obtained and describes the corrective actions taken in an effort to remedy those problems as the Reasoning Mind Project is implemented during the 2006-2007 school year. Finally, the report lists some questions needing to be addressed during the Project's implementation during the 2006-2007 school year and subsequent implementations of the Reasoning Mind Program.

Results Obtained from the Pilot Project

During the Spring 2003 semester, the Reasoning Mind Program was piloted at Hogg Middle School in the Houston Independent School District. An evaluation of that pilot program was conducted. That evaluation was designed to: (1) compare the mathematics achievement of seventh grade students who participated in the Reasoning Mind Pilot Project to the mathematics achievement of seventh grade students who did not participate in the Reasoning Mind Pilot Program; and (2) describe the attitudes of those students who participated in the Reasoning Mind Pilot Program regarding the Reasoning Mind instructional system they had experienced; and (3) to describe the attitudes of their teachers regarding the Reasoning Mind instructional system.

The results obtained regarding student achievement—whether measured by a pretest and a posttest designed for this purpose or the mathematics section of the state-mandated and District administered achievement test, the Texas Assessment of Knowledge and Skills—indicated that the mathematics achievement of students who participated in the Reasoning Mind Pilot Program—the experimental group—was statistically significantly higher than the mathematics achievement of students who did not participate in the Reasoning Mind Pilot Program—the control group. These results also indicated that the difference between the mathematics achievement of the experimental and control groups were also educationally meaningful; that is, they were of practical significance. In educational research, effect size—a standardized mean difference—is the statistic used to determine whether the difference between the means of two groups is of practical significance. An effect size equivalent to or greater than one-third of a standard deviation unit (+0.33) is considered to be of practical significance; that is, the difference is educationally meaningful. The analyses reported in this evaluation yielded effect size differences favoring the experimental group that were: +1.66, +1.64, +0.80, +1.61, +0.79, and +.048. This is clear evidence that participation in the Reasoning Mind Pilot Program made differences that were extraordinary—differences far beyond what would be reasonable given the focus and duration of the project.

The results obtained regarding student attitudes indicated that—for the most—the students who participated in the Reasoning Mind Pilot Program were very positive about their experiences. For example, 80.00% of those students reported that they liked learning mathematics on the Reasoning Mind system, only 6.67% reported that they preferred learning mathematics in their regular classes, many reported that the Reasoning Mind system was fun, and several reported that they hoped to be able to participate in the system again next year.

The results obtained regarding teacher attitudes indicated that the two teachers of the students who participated in the Reasoning Mind Pilot Program were very positive about the Reasoning Mind system. Both teachers reported that the system kept their students engaged and motivated and that it was effective in helping their students learn mathematics.

By any standard, the results obtained in this study provided solid evidence that the Reasoning Mind Pilot Program was highly effective. Therefore, when the Reasoning Mind Program was implemented in selected schools during the 2005-2006 school year, there was every expectation that the results would be positive. That is, the mathematics achievement of students who participated in the Reasoning Mind Program would be statistically significantly higher than the mathematics achievement of students who did not participate in the Reasoning Mind Program.

The Evaluation Plan for the 2006-2007 Project

In each of the schools in which the Reasoning Mind Project was implemented during the 2005-2006 school year, the intent was that the evaluation would utilize a pretest-posttest control group research design with random selection to examine the effectiveness of the Reasoning Mind program with regard to student achievement in mathematics. In effect, the evaluation plan called for a true experimental research design in which the independent variable was method of mathematics instruction (either student participation in the Reasoning Mind program or in traditional mathematics instruction) and the dependent variable was student achievement.

The pretest-posttest control group research design with randomization includes three features that make it a very powerful evaluation tool: (1) the use of a pretest; (2) the use of a control group; and (3) the random selection of students in the experimental and control groups. This research design allows posttest differences between the mathematics achievement of the experimental group—the students who participated in the Reasoning Mind program—and the mathematics achievement of the control group—the students who did not participate in the Reasoning Mind program—to be attributed to the experimental treatment—the Reasoning Mind program. Thus, the pretest-posttest control group with randomization is an excellent research design—the gold standard—for establishing cause and effect relationships as was the intent in this field test of the Reasoning Mind program.

Unfortunately, as is most often the case in field-based research involving curricular and instructional interventions in public school settings, efforts to utilize the pretest-posttest control group research design with randomization were compromised by a host of “real world” problems far too numerous to discuss here. However, the discussion of a few serious examples might be useful: (1) there was great difficulty in getting “good” data; for example, the principal of one of the schools refused to provide student mathematics achievement data while the principals at two other schools refused to administer the survey of student attitudes; (2) usually, principals designated “intact classes” of students to participate in the Reasoning Mind program or the control group (random assignment, not random selection) and the groups assigned were rarely comparable with regard to mathematics achievement; indeed, one principal assigned the school’s lowest performing students to the experimental group thinking it would help them most; (3) the ten schools that participated in the Project are schools that serve low-income, minority students and, as is often the case in such schools, there was high student mobility; this had two negative impacts on the evaluation; first, mathematics achievement data from the previous school year—the “pretest” data—were not available for a considerable number of students and, secondly, quite a few students moved from their schools during the year and, therefore, they were dropped from the evaluation and their data were excluded from the analyses; and (4) for a variety of reasons, in most instances, the Reasoning Mind program was not implemented as intended and, in many cases, was not used with the students for whom it had been designed. An examination of the results reported here should take these sorts of problems into account as it is clear that each of the ten schools involved in the Project presented a unique set of problems with regard to the implementation of the Reasoning Mind program and—consequently—with regard to the implementation of the external evaluation.

In each instance in which pretest and posttest student mathematics achievement data were available and in which an experimental group and a control (or comparison) group were identified, the evaluation utilized analysis of covariance procedures in which the independent variable was method of mathematics instruction—participation or non-participation in the Reasoning Mind program, the covariate was student mathematics achievement as measured by the pretest, and the dependent variables was student mathematics achievement as measured by the mathematics section of the Texas Assessment of Knowledge and Skills or the mathematics section of the Stanford Achievement Test Series. That is, analysis of covariance procedures were used to compare the posttest performance of each set of experimental and comparison groups. As noted earlier, the use of pretests and analysis of covariance procedures allows posttest differences to be attributed to the experimental treatment—participation in the Reasoning Mind program—rather than to initial differences between the groups and the use of control/comparison groups allows posttest differences to be attributed to the experimental treatment rather than to extraneous variables.

In those instances in which pretest and posttest student mathematics achievement were available for an experimental group but a comparison group was not identified, the evaluation utilized a pretest-posttest one-group only research design and *t*-tests for paired samples to compare the posttest performance of the experimental group to its pretest performance. Attributing posttest differences to the experimental treatment is done with caution when using this research design as it does not control for the possibility of extraneous variables.

In all statistical analyses, the ninety-five percent confidence level ($p \leq .05$) was used as the criterion for determining whether differences between means are statistically significant and an effect size—a standardized mean difference—equivalent to one-third of a standard deviation ($d \geq +0.33$) was used as the criterion for determining whether differences between means are educationally meaningful. These results are summarized in this report in Tables 1 and 2.

The results described in Tables 1, 2, and 3 prompted the completion of several post hoc analyses that were intended to provide a better understanding of those results. The first analyses utilized bivariate and multiple regression techniques to examine the relationships between a number of predictor variables and a criterion variable (the mathematics achievement of students in the experimental group as measured by the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2006). An additional analysis utilized discriminant function analysis to examine the mathematics achievement of students in the experimental group and identify the extent to which a group of six variables were predictive of a student making pretest-posttest gains or losses on the mathematics sections of the Texas Assessment of Knowledge and Skills administered in April 2005 and in April 2006.

In addition to focusing on student mathematics achievement, the evaluation also was designed to describe student perceptions regarding aspects of the Reasoning Mind program. A survey of student attitudes was made through the administration of a questionnaire and narrative and tabular descriptions of the results are summarized in Tables 8 through 11 in this report.

Student Mathematics Achievement: Results from Primary Analyses

Table 1 on the next page—page 5—presents the results that were obtained when pretest-posttest control group research designs and analysis of covariance procedures were used to compare the mathematics achievement of the experimental groups and the comparison groups for which mathematics achievement data were available. The results include the number of students in each group, the means, and standard deviations obtained for each group, and the F ratio, p value, and d statistic obtained for each comparison presented by school, grade level, and test—the mathematics section of the Texas Assessment of Skills and Knowledge (TAKS) or the mathematics section of the Stanford Achievement Test Series (SATS). The results from the Texas Assessment of Skills and Knowledge are presented as scale scores; except as noted, the results from the Stanford Achievement Test Series are presented as grade equivalent scores.

Table 2 on page 6 presents the results that were obtained when pretest-posttest one group only designs and t-tests for paired samples were used to compare the pretest and posttest mathematics achievement of the experimental groups in those instances where a comparison group was not designated. The results include the number of students in each group, the pretest and posttest means and standard deviations obtained for each group, and the F ratio, p value, and d statistic obtained for each comparison presented by school, grade level, and test—the mathematics section of the Texas Assessment of Skills and Knowledge (TAKS) or mathematics section of the Stanford Achievement Test Series (SATS). Except as indicated all scores are reported as raw scores—the number of items the student answered correctly—for all of the items in the mathematics section of the test.

Table 1

Results Obtained from Experimental Group-Comparison Group Comparisons of Mathematics Achievement as Measured by TAKS and SATS using Analysis of Covariance Procedures

			Experimental Group			Comparison Group					
School	Test	Grade	*N*	Mean	*SD*	*N*	Mean	*SD*	*F*	*p*	*d*
B	TAKS	5	14	2193.23	178.11	11	2167.44	150.90	0.28	.600	+0.16
F	TAKS	5	13	2237.46	233.19	12	2285.00	256.95	0.50	.485	-0.19
A	TAKS	6	30	2069.73	182.88	21	2100.33	144.48	0.81	.372	-0.19
B	TAKS	6	18	2036.08	127.00	15	2090.31	172.56	1.99	.168	-0.36
C	TAKS	6	10	2012.30	136.43	9	2079.67	160.34	1.84	.194	-0.45
E	TAKS	6	33	2001.80	142.73	26	1991.14	141.64	0.10	.755	+0.07
F	TAKS	6	14	2239.82	162.19	15	2286.44	152.11	0.85	.366	-0.30
G	TAKS	6	25	2027.43	191.58	23	2059.62	144.84	0.83	.367	-0.19
H	TAKS	6	6	2027.94	158.62	12	2000.78	83.50	0.82	.381	+0.21
I	TAKS	6	69	2252.82	156.48	74	2230.65	180.83	1.13	.290	+0.13
A	TAKS	7	47	2035.26	97.13	46	2059.00	95.82	1.83	.179	-0.25
B	TAKS	7	22	2059.19	87.65	22	2076.86	95.56	0.52	.475	-0.19
E	TAKS	7	25	2032.16	87.27	29	2033.38	91.19	0.00	.950	-0.01
G	TAKS	7	32	2024.90	68.23	21	2004.20	144.84	1.34	.253	+0.18
All	TAKS	5-7	443	2106.99	165.11	336	2109.80	164.23	0.10	.747	-0.02
B	SATS	5	14	6.12	1.33	20	1.63	5.45	2.41	.131	+0.45
B	SATS	6	19	5.69	1.52	16	6.21	1.48	1.14	.294	-0.34
C	SATS	6	10	3.04	1.53	9	6.45	1.82	0.40	.534	-0.24
E	SATS	6	31	5.98	1.19	27	6.25	1.26	0.98	.327	-0.22
G	SATSpr[1]	6	22	9.23	4.09	20	6.80	3.43	2.07	.045	+0.64
G	SATSps[2]	6	22	17.96	5.84	20	17.50	5.82	0.25	.802	+0.08
H	SATS	6	15	6.52	1.74	12	5.87	1.32	2.10	.161	+0.39
A	SATS	7	46	7.12	1.61	46	7.49	1.92	1.21	.274	-0.27
B	SATS	7	23	6.52	1.85	22	6.41	1.18	0.08	.778	+0.07
E	SATS	7	25	6.78	1.49	28	7.25	1.68	1.80	.185	-0.29
G	SATSpr[1]	7	23	8.26	4.99	14	8.64	3.95	0.24	.809	-0.08
G	SATSps[2]	7	23	19.00	7.03	14	18.50	6.26	0.22	.828	+0.08

[1]Scores for the Procedures Section of the SATS rather than total scores; scores are raw scores.
[2]Scores for the Problem Solving Section of the SATS rather than total scores; scores are raw scores.

Table 2

Results Obtained from Pretest-Posttest Comparisons of Mathematics Achievement as Measured by TAKS and SATS using t-Tests for Paired Samples

				Pretest		Posttest				
School	Test	Grade	*N*	Mean	*SD*	Mean	*SD*	*t*	*p*	*d*
D	TAKS	6	36	2176.03	173.34	2183.68	165.50	0.23	.813	+0.05
C	TAKS	7	20	1955.10	119.23	2035.40	80.50	3.91	.001	+0.79
D	TAKS	8	27	2050.07	94.05	2160.30	82.01	6.23	<.001	+1.25
C	SATS	7	20	5.87	1.03	7.46	1.79	4.54	<.001	+1.09

At first glance, the results reported in Tables 1 and 2 might be described as somewhat disappointing in that only four of the thirty comparisons yielded statistically significant differences. However, all four of those favored the experimental group and—with effect sizes greater than +0.33—all four were also educationally meaningful. While the Reasoning Mind program did not appear to be as effective as had been hoped, there are two additional comments that need to be made at this point. The first of these has been made previously—the implementation and evaluation of the Reasoning Mind program were fraught with problems. As noted earlier, many of those were the kinds of problems that are inherent in field testing a technology-dependent curricular and instructional intervention in a public school setting. Many of those problems could not have been anticipated and avoided. This field test has provided valuable experience and useful information concerning those problems that might have been avoided. In that regard, the evaluation should be viewed as formative as it has provided the developers with important information about how to better proceed. In short, this field test did just what a field test is supposed to do—lead to the identification of problems—and potential solutions—so that improvements can be made.

The second comment that needs to be made is that the results may be more positive than they seem on first examination. When meta-analysis—a technique in which statistical procedures are used to identify trends in the statistical results yielded by a set of studies concerning the same research problem—were used to examine the results obtained from the thirty comparisons reported in Tables 1 and 2, a positive effect size—a mean effect size of +0.07—was obtained. However, as shown in Table 1, when analysis of covariance procedures were used to compare the posttest mathematics achievement of all students in the experimental groups and all students in the comparison groups for whom there was also pretest mathematics achievement data, the adjusted mean obtained for the comparison groups (N = 336, mean = 2109.80) was slightly higher than the adjusted mean obtained for the experimental groups (N = 443, mean = 2106.99). This slight difference was not statistically significant (p = .747) and the effect size (d = -0.02) was very small. At the least, these results suggest that the program did no harm. This finding should encourage the program's developers to seek ways to continue to improve the curriculum and its implementation and evaluation.

In addition to yielding a mathematics scale score as an indicator of a student's performance on the mathematics section of the Texas Assessment of Knowledge and Skills, the Texas Education Agency indicates whether that student has "met standards," that he or she has, in effect, passed the test. Therefore, it made sense to examine the compare use these scores in comparing the performance of the students in the experimental and comparison groups. Specifically, McNemar's chi square technique was used to compare the percentage of students in the experimental and comparison groups who met standard on the administrations of the mathematics section of the Texas Assessment of Knowledge and Skills—the tests administered in April 2005 and April 2006. The results are presented in Table 3 below.

Table 3

Results Obtained from Experimental Group-Comparison Group Comparisons of Percentage of Students who "Met Standard" in Mathematics as Indicated by Pretest and Posttest Administrations of TAKS Using McNemar's Chi Square Technique

Group	Pretest	Posttest
Experimental	39.10%	50.11%
Comparison	37.09%	45.46%

$\chi^2 = 2.00$, $p = .157$

As shown in Table 3, when the percentage of students in the experimental and control groups who met standard on the April 2005 administration of the mathematics section of the Texas Assessment of Knowledge and Skills (the pretest) was compared to the percentage of students in the experimental and control groups who met standard on the April 2006 administration of the mathematics section of the Texas Assessment of Knowledge and Skills (the posttest), the analysis yielded a chi square statistics ($\chi^2 = 2.00$) that was not statistically significant ($p = .157$) even though it showed that the percentage of students in the experimental groups who met standard on the posttest (50.11%) had increased by 11.01% from the pretest (39.11%) while the percentage of students in the control groups who met standard on the posttest (45.46%) had increased by only 8.37% from the pretest (37.09%). Nonetheless, these results are encouraging.

Student Mathematics Achievement: Results from Post Hoc Analyses

The results described in Tables 1, 2, and 3 prompted three sets of post hoc analyses that were intended to provide a better understanding of those results. The sections that follow describe those analyses and present the results obtained.

Bivariate Regression Analyses. The first set of analyses utilized bivariate regression techniques—the Pearson product-moment correlation technique—to examine the relationship between the criterion variable (mathematics achievement gain or loss—the increase or decrease in student mathematics achievement of each student in the experimental group as indicated by the difference between the student's scale score on the mathematics section of the Texas

Assessment of Knowledge and Skills administered in April 2005—the pretest—and the student's scale score on the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2006—the posttest) and each of five predictor variables: (1) the number of hours each student was "logged" into the Reasoning Mind program; (2) the quality of the administrative support in each of the host schools as rated by program staff using a ten-point scale in which "1" was "low" and "10" was "high" and weighted by a factor of one; (3) the quality of the classroom facilities in each classroom in which the Reasoning Mind was implemented as rated by program staff using a ten-point scale in which "1" was "low" and "10" was "high" and weighted by a factor of three; (4) the quality of each of the teachers who utilized the Reasoning Mind program as rated by program staff using a ten-point scale in which "1" was "low" and "10" was "high" and weighted by a factor of six; and (5) the reading achievement of each student in the experimental group as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006).

The purpose of each of these analyses was to describe the relationship between student mathematics achievement gains (and losses) and one of those five predictor variables. For example, because use of the Reasoning Mind program requires that the student read and understand both content and directions, it seemed reasonable to examine the nature of the relationship between the mathematics achievement and reading achievement of the students who participated in the Reasoning Mind program. That is, to what extent did "better" readers tend to make greater gains from the pretest to the posttest than did "poorer" readers? The results obtained from these analyses are presented in Table 4 below.

Table 4

Results Obtained from Examination of Relationships Between Selected Predictor Variables and Student Mathematics Achievement Gain Scores Using the Person Product Moment Correlation Technique

Variable	*N*	Mean	*SD*	*r*	*p*
Hours Student was "Logged On"	443	56.83	35.47	-0.06	.205
Quality of Administrative Support	443	6.04	3.08	-0.01	.851
Quality of Classroom Facilities	443	18.31	3.42	+0.00	.937
Quality of Teachers	443	30.61	14.78	+0.06	.247
Reading Achievement	357	2129.29	152.37	+0.07	.192

As shown in Table 4, none of the analyses yield a statistically significant correlation coefficient. Indeed, the correlation coefficients obtained indicate that in no case was there a relationship between a predictor variables and the criterion variable. Thus, the following interpretation is warranted: There are no statistically significant relationships between the mathematics achievement gains or losses of the students who participated in the Reasoning

Mind program and the number of hours each student was "logged" into the Reasoning Mind program, the quality of the administrative support in the host schools, the quality of the classroom facilities in each classroom in which the Reasoning Mind was implemented, the quality of each of the teachers who utilized the Reasoning Mind program, and the reading achievement of the students as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006.

Multiple Regression Analyses. In spite of the results obtained from the bivariate regression analyses, the multiple regression technique was used to examine the relationship between a combination of those five predictor variables (the number of hours each student was "logged" into the Reasoning Mind program, the quality of the administrative support in each of the host schools, the quality of the classroom facilities in each classroom in which the Reasoning Mind was implemented, and the quality of each of the teachers who utilized the Reasoning Mind program, and the reading achievement of each student in the experimental group as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006) and the criterion variable (the mathematics achievement of students in the experimental groups as indicated by posttest mathematics achievement scores and as indicated by pretest-posttest gain scores). These analyses were undertaken because, for example, it is reasonable to expect that a "better" reader who was taught by a "better" teacher would make greater gains in mathematics achievement than a "poorer" reader who was taught by a "poorer" teacher.

First, the multiple regression technique using the backward stepwise process was used to determine what combination of these five variables are the best predictors of student mathematics achievement as measured by the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2006 (the posttest). The results from that analysis are presented in Table 5 below. Next, the multiple regression technique using the backward stepwise process was used to determine what combination of these five variables are the best predictors of student mathematics achievement as indicated by the pretest-posttest gain scores—the difference between the scale scores yielded by the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2005 (the pretest) and the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2006 (the posttest). The results obtained from this analysis are presented in Table 6 on the next page.

Table 5

Results Obtained from Backward Stepwise Multiple Regression Used to Identify Variables that are Predictive of Mathematics Achievement of Students in Experimental Groups as Indicated by Posttest Scale Scores

Dependent Variable: Posttest Scale Scores	N: 357
Multiple R: .61	Squared Multiple R: .37
Adjusted Squared Multiple R: .36	Standard Error of Estimate: 121.73

Variable	B	Standard Error	β	Tolerance	*t*	*p*
Constant	1139.15	96.40	.00		11.82	<.001
SchAdm	-9.88	2.23	-.19	.94	4.44	<.001
RMHours	-0.66	0.17	-.17	.97	3.92	<.001
Reading	+0.49	0.04	+.49	.95	11.18	<.001

Analysis of Variance

Source	Sum of Squares	*df*	Mean Square	*F*	*p*
Regression	3022352.21	3	1007450.74	67.98	<.001
Residual	5231197.78	353	82.00		

Table 6

Results Obtained from Backward Stepwise Multiple Regression Used to Identify Variables that are Predictive of Mathematics Achievement of Students in Experimental Groups as Indicated by Pretest-Posttest Gain Scores

Dependent Variable: Pretest-Posttest Gain Scores	N: 357
Multiple R: .18	Squared Multiple R: .03
Adjusted Squared Multiple R: .03	Standard Error of Estimate: 133.04

Variable	B	Standard Error	β	Tolerance	*t*	*p*
Constant	-2893.12	100.19			2.89	.004
Teacher	+0.78	0.48	+.09	.99	1.63	.104
Reading	+0.15	0.05	+.17	.99	3.15	.002

Analysis of Variance

Source	Sum of Squares	*df*	Mean Square	*F*	*p*
Regression	218044.95	2	109022.47	6.16	.002
Residual	6265718.30	354	17699.77		

The objective in using a backward stepwise multiple regression is to identify which combination of a set of predictor variables is most predictive of a criterion variable. In the analysis that yielded the results reported in Table 5, the criterion variable was the mathematics achievement of the students in the experimental groups as indicated by the scale scores on the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2006. As shown in Table 5, the backward step multiple regression yielded a squared multiple *R* (a coefficient of determination) which was .37. This indicates that there was a very weak association between the criterion variable (mathematics achievement as measured by the posttest) and the combination of three of the five predictor variables as indicated by their standardized regression coefficients (β weights). Those three variables were: (1) the quality of the administrative support in the host schools (β = -.19); (2) the number of hours each student was "logged" into the Reasoning Mind program (β = -.17); and (3) the reading achievement of the students as measured by the reading section of the Texas Assessment of Knowledge and

Skills administered in April 2006 ($\beta = +.49$). In short, student achievement in mathematics was only moderately associated with the three predictor variables identified by this analysis and, of the three, most of the variance was accounted for by reading achievement.

In the analysis that yielded the results reported in Table 6, the criterion variable was the mathematics achievement of the students in the experimental groups as indicated pretest-posttest gain scores. As shown in Table 6, the backward step multiple regression yielded a squared multiple *R* (a coefficient of determination) which was only .03. This indicates that there was nearly no association between the criterion variable (mathematics achievement as indicated by pretest-posttest gain scores) and the combination of two of the five predictor variables as indicated by their standardized regression coefficients (β weights). Those two variables were: (1) the quality of the teacher who taught the students ($\beta = +.09$); and (2) the reading achievement of the students as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006 ($\beta = +.17$). As with the first of these two analyses, student achievement in mathematics as indicated by pretest-posttest differences was only slightly associated with the two predictor variables identified and, as previously, most of the variance was accounted for by reading achievement.

Discriminant Function Analysis. The final post hoc analysis utilized a procedure called discriminant function analysis to determine what variables—if any—discriminate between those students in the experimental group who "made pretest-posttest gains" and those who "made pretest-posttest loses" in mathematics achievement as measured by the mathematics sections of the Texas Assessment of Knowledge and Skills administered in April 2005 and in April 2006. This analysis examined six predictor variables (the number of hours each student was "logged" into the Reasoning Mind program, the quality of the administrative support in each of the host schools, the quality of the classroom facilities in each classroom in which the Reasoning Mind was implemented, and the quality of each of the teachers who utilized the Reasoning Mind program, the mathematics achievement of each student in the experimental group as measured by the mathematics section of the Texas Assessment of Knowledge and Skills administered in April 2005, and the reading achievement of each student in the experimental group as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006). More specifically, a backward stepwise discriminant function analysis was used in this analysis to examine the extent to which those six predictor variables would correctly classify students as having "made pretest-posttest gains" or as "having made pretest-posttest loses."

The purpose of this analysis was to determine to what extent which of these six variables—in combination—would most accurately predict group membership. As an example, if there are two groups of teachers who are, by some accepted definition, elementary school teachers (defined as teachers who teach in pre-kindergarten through fifth grade) and secondary school teachers (defined as teachers who teach in sixth grade through twelfth grade), discriminant function analysis procedures should correctly classify 100% of the teachers in a sample as either being an elementary school teacher or an secondary school teacher if each teacher teaches only in kindergarten through fifth grade or teaches only sixth grade through twelfth grade and the grade level or levels taught by each teacher is known and entered into the calculation.

In the analysis reported here, the intent was to attempt to "classify" students into two groups—those students who made pretest-posttest gains in mathematics achievement and those students who made pretest-posttest losses in mathematics achievement. The results obtained from this analysis are presented in Table 7

Table 7

Results Obtained from Backward Stepwise Discriminant Function Analysis Which Classified Students in the Experimental Groups into Groups in Which Students Made Mathematics Achievement Gains and in Which Students Made Mathematics Achievement Using Three of the Six Predictor Variables

	Predicted Group Membership		
Group	Made Gains	Made Loses	Total
Made Gains	79 (65.83%)	41 (34.17%)	120 (100.00%)
Made Losses	55 (23.21%)	182 (76.79%)	237 (100.00%)

73.11% of students (261 of 357) were correctly classified.

The results presented in Table 7 indicate that an analysis of the predictor variables yielded classifications that were correct 73.11% of the time. The analysis also indicated that: (1) three of the six predictor variables examined—quality of the administrative support in each of the host schools, the mathematics achievement of the student as measured by the Texas Assessment of Knowledge and Skills administered in April 2006, and the reading achievement of the student as measured by the reading section of the Texas Assessment of Knowledge and Skills administered in April 2006—were the predicator variables that—in combination—yielded the most accurate classification; and (2) those three variables were somewhat more able to predict which students made losses (76.79% percent accuracy) than they were to predict which students made gains (65.83%). These results provide some additional insight into the results obtained in this evaluation.

Student Attitudes

While the major focus of this evaluation was to describe the effect of the Reasoning Mind program on student achievement in mathematics, a secondary focus was on describing how students felt about several aspects of the program. In those schools were it was allowed, a survey of student attitudes was administered to those students who had participated in the Reasoning Mind program—the experimental group—toward the end of the school year. The discussion below describes the results obtained from the four scales contained in that survey.

The first item asked students to select one of four responses to the following question: "Did you like learning math on the RM system?" The four alternatives give were: "I liked it a whole lot" (4); "I liked it" (3); "It was O.K. (2); and "I didn't like it at all" (1). The results obtained are presented in Table 7 on the following page.

Table 8

Results Obtained from Analysis of Student Responses to Item 1 on Student Survey: "Did you like learning math on the RM system?"

School		1	2	3	4	Total	Mean	*SD*
A	*N*	13	48	10	4	75	2.07	0.72
	%	17.33	64.00	13.33	5.33	100.00		
B	*N*	0	2	2	7	11	3.46	0.82
	%	0.00	18.18	18.18	63.64	100.00		
C	*N*	4	16	7	1	28	2.18	0.72
	%	14.29	57.14	25.00	3.57	100.00		
E	*N*	4	20	12	2	38	2.32	0.74
	%	10.53	52.63	31.58	5.26	100.00		
G	*N*	8	34	10	5	57	2.21	0.80
	%	14.04	59.65	17.54	8.77	100.00		
H	*N*	2	5	4	2	13	2.46	0.97
	%	15.39	38.46	30.77	15.39	100.00		
I	*N*	13	59	26	38	136	2.65	0.99
	%	9.56	43.38	19.12	27.94	100.00		
J	*N*	1	5	5	1	12	2.50	0.80
	%	8.33	41.67	41.67	8.33	100.00		
Total	*N*	45	189	76	60	370	2.41	0.90
	%	12.16	51.08	20.54	16.22	100.00		

As shown in Table 8, on a four-point scale in which "4" was "I liked it a whole lot" and "1" was "I didn't like it at all," a bit more than half of all students (51.08%) reported that "It was O.K." while another fifth (20.54%) reported that they "liked it." While this may not be as positive a response as might have been hoped for, it should be noted that—overall—only 12.16 reported that they "didn't like it at all."

The second item on the survey asked the student "In the future, where would you like to learn math most?" and gave the student the following three choices: "On the RM website," "In my regular math class," and "It doesn't matter." Table 9 on the next page presents the results obtained for Item 2 of the student survey.

Table 9

Results Obtained from Analysis of Student Responses to Item 2 on Student Survey: "In the future, where would you like to learn math most?"

School		RM Website	Regular Class	Doesn't Matter	Total
A	*N*	24	16	35	75
	%	32.00	21.33	46.67	100.00
B	*N*	5	3	3	11
	%	45.46	27.27	27.27	100.00
C	*N*	6	5	17	28
	%	21.43	17.86	60.71	100.00
E	*N*	14	9	15	38
	%	36.84	23.68	39.47	100.00
G	*N*	15	22	20	57
	%	26.32	38.60	35.09	100.00
H	*N*	7	3	3	13
	%	53.85	23.08	23.08	100.00
I	*N*	64	31	41	136
	%	47.06	22.79	30.15	100.00
J	*N*	2	4	6	12
	%	16.67	33.33	50.00	100.00
Total	*N*	137	93	139	370
	%	37.03	25.14	37.84	100

As shown in Table 9, there was variation among schools while—overall—only a bit more than one-fourth of the students (25.24%) reported that—in the future—they would prefer learning mathematics in their regular classrooms. Further, the results indicate that better than a third of all of the students indicated that they would prefer learning mathematics on the RM website (37.03%) while another third or so indicated that they had no preference (37.84%). Thus, it seems reasonable to suggest that about three-fourths of the students who responded to the survey would be comfortable learning mathematics on the Reasoning Mind website.

The third item on the survey asked the student "Which class was harder this year?" and gave the student the following five choices: "The RM class was much harder than my regular math class," "The RM class was harder than my regular math class," "They were about the same," "My regular math class was harder than RM" and "My regular math class was much harder than RM." Table 10 presents the results obtained for Item 3 of the student survey.

Table 10

Results Obtained from Analysis of Student Responses to Item 3 on Student Survey: "Which class was harder this year?"

School		RM Much Harder	RM Harder	About the Same	Regular Class Harder	Regular Class Much Harder	Total
A	*N*	2	3	41	18	11	75
	%	2.67	4.00	54.67	24.00	14.67	100.00
B	*N*	3	2	2	2	1	10
	%	30.00	20.00	20.00	20.00	10.00	100.00
C	*N*	4	5	9	8	2	28
	%	14.29	17.86	32.14	28.57	7.14	100.00
E	*N*	0	3	18	14	3	38
	%	0.00	7.90	47.37	36.84	7.90	100.00
G	*N*	0	6	28	18	5	57
	%	0.00	10.53	49.12	31.58	8.77	100.00
H	*N*	0	0	7	3	3	13
	%	0.00	0.00	53.85	23.08	23.08	100.00
I	*N*	21	20	55	31	9	136
	%	15.44	14.71	40.44	22.79	6.62	100.00
J	*N*	1	0	6	3	2	12
	%	8.33	0.00	50.00	25.00	16.67	100.00
Total	*N*	31	39	166	97	36	369
	%	8.40	10.57	44.98	26.29	9.76	100

As shown in Table 10, a majority of the students reported that their regular mathematics class was "about the same" (44.98%) or "harder" (26.29%) than RM. Less than ten percent of the students reported that their RM class was "much harder" (8.40%) and less than ten percent of the students reported that their regular class was "much harder" (9.76%). This seems to confirm the earlier finding that a large majority of the students would be comfortable learning mathematics in the RM system.

The fourth item—the last of the scales on the survey—asked the student "Where have you learned more math this year?" and gave the student the following five choices: "I learned much more in RM than in my regular math class," "I learned more in RM than in my regular math class," "I learned about the same," "I learned more in my regular math class than in RM," and "I learned much more in my regular math class than in RM." Table 11 presents the results obtained for Item 4 of the student survey.

Table 11

Results Obtained from Analysis of Student Responses to Item 4 on Student Survey: "Where have you learned more math this year?"

School		Much More in RM	More in RM	About the Same	More in Regular Class	Much More in Regular Class	Total
A	*N*	12	9	32	12	9	74
	%	16.22	12.16	43.24	16.22	12.16	100.00
B	*N*	4	0	3	4	0	11
	%	36.36	0.00	27.27	36.36	0.00	100.00
C	*N*	8	3	12	1	4	28
	%	28.57	10.71	42.86	3.57	14.29	100.00
E	*N*	6	11	14	7	0	38
	%	15.79	28.95	36.84	18.42	0.00	100.00
G	*N*	9	5	22	11	10	57
	%	15.79	8.77	38.60	19.30	17.54	100.00
H	*N*	1	2	7	1	2	13
	%	7.69	15.39	53.85	7.69	15.39	100.00
I	*N*	27	21	35	28	25	136
	%	19.85	15.44	25.74	20.59	18.38	100.00
J	*N*	4	0	2	3	3	12
	%	33.33	0.00	16.67	25.00	25.00	100.00
Total	*N*	71	51	127	67	53	369
	%	19.24	13.82	34.42	18.16	14.36	100

As shown in Table 11, about one-third of the students (33.06%) reported that they had learned "much more" (19.24%) or "more" (13.82%) mathematics in the RM class than in their regular class, about one-third of the students (34.4%) reported they learned "about the same" in the RM and regular classes, and about one-third (32.52%) reported that they had learned "much more" (14.36%) or "more" (18.16%) in their regular class than in the RM class. This seems to suggest that—at a minimum—the Reasoning Mind did "no harm" in the minds of the students who responded to the survey.

Summary of the Results

To this point, this report has described the results obtained from an evaluation of the Reasoning Mind Project—a project intended to enhance student achievement in mathematics which was field tested in ten schools in the Houston metropolitan area during the 2005-2006 school year. First, the report provided a brief description of the evaluation plan. Next, the report presented a summary of the results obtained from analyses of student mathematics achievement data. Those results suggested that a well-implemented Reasoning Mind program may have the potential to yield positive results regarding student mathematics achievement. At the least, it is totally reasonable to argue that the program did no harm.

Further analyses are undertaken in an effort to better understand what worked and what did not work—and why. Unfortunately, most of those post hoc analyses were not particularly successful in identifying variables that were strongly predictive of student mathematics achievement. Not surprisingly, the results suggest that student achievement in reading and student performance on prior measures of mathematics achievement are the best predictors of success in mathematics. Finally, the report provided a narrative and tabular summary of the results obtained from analyses of student attitudes toward the Project. In the main, students seem to like—or not dislike—learning mathematics in the Reasoning Mind class.

Some Observations

The view here is that a meaningful evaluation of a program such as the Reasoning Mind Program must implement an evaluation plan that embodies the characteristics of rigorous experimental research design to the extent possible in a field setting. In such an evaluation plan, the program—the instructional intervention—is seen as the "experimental treatment." Thus, if the evaluation is to yield valid information concerning the effect of the program, it is absolutely critical that there be "treatment fidelity." That is, the experimental treatment—in this case, the Reasoning Mind program—must be implemented as intended. Anything less jeopardizes the validity of the findings.

A recent publication, *Technology in Schools: What the Research Says* (Metiri Group, 2006) provides an insightful discussion of the problems associated with attempts to use technology in schools to enhance student learning. The authors argue that "**overall, across all uses in all content areas, technology does provide a small, but significant, increase in learning when implemented with fidelity**" (Metiri Group, 2006, page 15, emphasis by the authors). They add: "The indicators for success are not solely dependent on the level of student access, but rather on the nature of the student and teacher use and the fidelity of the implementation. **Such fidelity of implementation in a school, in turn, is determined by leadership, teacher proficiency, professional development, fit with curriculum, school culture, pedagogical approaches—and to some degree on levels and types of technology access"** (Metiri Group, 2006, page 16, emphasis by the authors).

The authors also note that "despite the decades of use of technology in elementary and secondary schools, the number of rigorous research studies is small, the quality of the studies varies considerably, and the level of funding for such research is low in most countries (Metiri Group, 2006, page 16). Earlier in their discussion, the authors define "rigorous research" as experimental or quasi-experimental design studies (*i.e.*, use of treatment and control groups, preferably through randomization, and rigorous statistical design and analysis to test hypotheses (Metiri Group, 2006, page 4). That was clearly the intent of this evaluation.

Unfortunately, in spite of the best intentions of those who developed and implemented the Reasoning Mind program and who commissioned this evaluation, because of treatment fidelity problems, this evaluation has produced little meaningful evidence concerning the program's effectiveness—or ineffectiveness. The evaluation did not provide convincing evidence that the program "worked" but neither did it provide evidence that it would not have "worked" had it been properly implemented. As noted earlier, the results do suggest that the program did no harm. Additionally, as noted in the next section of this report, it did reveal some of the problems that must be addressed as efforts to improve the program—and its evaluation—move forward.

Staff Perceptions

The Pilot Project conducted during the Spring 2003 semester at Hogg Middle School had been extraordinarily successful. The results obtained from the Project conducted during the 2005-2006 school year were disappointing. In an effort to gain a greater understanding of the problems that negatively affected these results, Reasoning Mind staff responsible for the design, development, implementation, and operation of the Reasoning Mind Project were asked to identify the most serious of those problems and to discuss the corrective actions taken in an effort to solve those problems as the Program is implemented during the 2006-2007 school year. The sections that follow discuss the problems identified and the corrective actions taken. The problems identified fall into two categories: (1) factors that are **directly controlled by the Reasoning Mind Project** (quality of curriculum, rich functionality of the system, and reliable performance of the software, for example); and (2) factors that are **directly controlled by schools and the school districts** (teacher competence and performance, program organization and administration, computer availability, reliability of school computer networks and connectivity to the Internet, for example). The ultimate success of the Reasoning Mind Project depends on how effective the organization is in improving factors in the first category and the extent to which Reasoning Mind staff can find means to indirectly control the factors from the second category.

Although significant mistakes were made during the Project's implementation during the 2005-2006 school year, as shown in the discussions which follow, most of them were successfully corrected in preparation to Project's implementation during the 2006-2007 school year. This is evidenced by how smoothly the Reasoning Mind Project has been running at schools since August 2006; this is in stark contrast to the previous school year. Reasoning Mind staff have also developed and implemented a business model for working with schools. The model prescribes strict guidelines for the program implementation at schools and is aimed at ensuring that these guidelines are followed. In short, these guidelines should enhance treatment fidelity.

Student Selection. In the Pilot Project at Hogg Middle School, students were selected from a group of 60, based on the results of a pre-qualification test. The weakest students in the class were excluded from this group, as they were not prepared to learn ratios and proportions through the Reasoning Mind system. This meant that students in both the experimental and control groups had higher than average mathematics and reading skills. In the 2005-2006 Project, students were selected from among the weakest students in the class—those who had the highest chance of failing to meet state requirements by passing TAKS. Those students had very weak mathematics and reading skills; in addition, many had low self-esteem, were not motivated to learn, and were frequently discipline problems.

In an effort to address the student selection problem, although the students participating in the 2006-2007 Project are fifth and sixth grade students (about 250 to 300 in each grade level), the Project is focused on fifth grade students who were selected at random from pools of "average" students; they are not the weakest students in their classes. Their level of preparedness for the Reasoning Mind curriculum is significantly higher than that of students in the 2005-2006

Project. The sixth grade students in the Project do come from among the weakest students in their classes, as was the case in last year. They have been included in the project to see if the Reasoning Mind program can be used as a supplemental remediation curriculum when implemented with fidelity.

Teachers. In the Hogg Middle School Project, Julia and Alex functioned as the teachers. Although they did not provide direct instruction to students, they were active in answering student questions. They also instilled an atmosphere of excitement and challenge. Students were continuously encouraged to challenge themselves; (for example, students were often told “you can do this” and “this is not that hard”). As a consequence, the confidence of students with regard to their ability to do mathematics increased dramatically Failure was not considered an option for any student in the class. Also, Julia and Alex had full ownership of the Project and had a strong desire to make it successful. In contrast, teachers in the 2005-2006 Project lacked training in both the Reasoning Mind curriculum and teaching method. They were often intimidated by the system since they had no knowledge about how the system guides students through the curriculum. Teachers were also intimidated by their inability to solve some of the more difficult problems offered to students. Their “buy-in prior” to the start of the Project was not ensured and teachers had no tangible incentives to make Project work.

Clearly, the Reasoning Mind system assumes that is the teacher’s responsibility to keep the students interested, to answer their questions, and ensure that each student works hard on the system. If the Reasoning Mind system does not give teachers the tools and training that they need to do their job, students will become bored, frustrated, and lazy. The fact that many students did not perform well in the 2005-2006 Project was in part due to the lack of active engagement and ownership on the part of teachers. Working with teachers to help them “take ownership” of the Reasoning Mind implementation in their classrooms is critical to the success of the entire initiative.

In an effort to deal with these issues, teachers in the 2006-2007 Reasoning Mind Project classrooms received an extensive five-day training in July 2006. Each was trained in the Reasoning Mind curriculum, classroom management, effective instructional strategies, and passed a final examination which assessed their mastery with regard to teaching with the Reasoning Ming system. By any standard, teachers are much better prepared for the 2006-2007 Project than they were in the 2005-2006 Project. Also, this year’s teachers enjoy much better support from Reasoning Mind personnel, as the position of Deployment Coordinator has been created expressly for this purpose. Reasoning Mind currently employs four Deployment Coordinators, each of whom is assigned to a certain number of teachers. The Deployment Coordinators are implementing newly created procedures for supporting and supervising teachers and are working closely with them to make sure that the implementation is going well. Teachers’ initial buy-in and “ownership” of the program was ensured during the summer training and is being continuously strengthened through their interaction with the Deployment Coordinators and through the accumulation of positive experiences using the program in their classrooms. Lastly, a special incentive program has been created for teachers; their work will be rewarded with performance-based bonuses at the end of the school year.

Curriculum Quality. The first release of the Reasoning Mind fifth grade curriculum had many problems; for example, some of the theory material presentations were confusing to students, many problems had “bugs” or overly wordy solutions, and so on. Solutions that are too long tended to be skipped by students. This leads students to waste time repeating mistakes rather than learning how to solve the problem. Since reading is difficult for many fifth graders, it is less likely that the students will make it through the curriculum if they must struggle through reading all of the long solutions.

One of the biggest mistakes in the 2005-2006 curriculum was the use of "composite" problems made up of four or five independent parts; when a student made a mistake in solving just one of the parts, the entire problem was counted as incorrect. The student was unable to easily see where he or she had made a mistake, and this led to increased frustration which negatively affected the student's self-esteem and motivation. The curriculum used in the Hogg Middle School program did not have nearly so many composite problems.

Another big mistake made in the 2005-2006 curriculum was to escalate the difficulty level of problems too quickly. For every problem students answered correctly, the Genie gave them a problem that was more difficult. Though effective for strong students, this method of giving problems may have been too rough for weaker students. This led to increased student frustration and disengagement from the Reasoning Mind curriculum.

In an effort to improve the curriculum, over the summer of 2006, the Reasoning Mind fifth grade curriculum was re-worked for improved quality. Those parts of the curriculum that had been identified as weakest based on the 2005-2006 experience, were completely re-written. Almost all "composite" problems were broken into separate problems. The pedagogy was adjusted to provide for a much smoother escalation of problem difficulty level. The resulting curriculum, as confirmed by students' experiences in during the first several months of the 2006-2007 Project, has become more understandable and significantly better. Students stay continuously engaged by the system and are challenged at the level of their abilities. However, further work is required to simplify and shorten solutions to many problems. This work is currently underway.

Curriculum Alignment. The curriculum used at Hogg Middle School was fully aligned with grade level expectations (ratios and proportions for sixth and seventh grade students), while most of the students in the 2005-2006 Project were sixth and seventh grade students who were studying Reasoning Minds fifth grade curriculum. The shortage of fifth graders in the 2005-2006 project prevented the ability to gain a solid understanding of the strengths and weaknesses of the Reasoning Mind system with the students for whom the program was designed.

To deal with this problem, the 2006-2007 Project includes many fifth grade students who are studying the Reasoning Mind fifth grade curriculum. This means that there is complete alignment of the curriculum with grade-level standards. In addition, the first part of the Reasoning Mind sixth grade curriculum will be deployed in January 2007, significantly enhancing curriculum alignment for the sixth grade students enrolled in the program.

Assessment. 2003 was the first year when the TAKS was administered. For this reason, students in the experimental and control groups could not and were not taught to the test since no one knew what the new test would be like. Thus, in 2003, TAKS was measuring student knowledge of mathematics—not how well their teachers had taught to the test. In the 2005-2006 Project, the achievement of students in the experimental groups—those who participated in the Reasoning Mind Project—was compared to the achievement of students in control groups—students who had spent a considerable amount of time preparing for the test.

This put the Reasoning Mind students at a disadvantage since the Reasoning Mind curriculum does not emphasize teaching to the test. Also, in the Hogg Middle School Project, student performance was measured not only on the TAKS, but also with a special test focusing on ratios and proportions developed by Reasoning Mind staff. No similar instrument was used to compare student mathematics skills in the 2005-2006 Project.

In an effort to deal with this problem, all students enrolled in the 2006-2007 Reasoning Mind Project have been pre-tested with a Reasoning Mind "entrance examination" and will be post-tested with an "exit examination." These tests have been designed to measure student knowledge of grade-level mathematics and the extent to which students are prepared for the next grade. Student performance on these examinations, in addition to their performance on TAKS, will be used to measure the Project's effectiveness.

System Performance. Although both the 2003-Hogg system and the 2005-2006 Project "crashed" often, the presence of Julia and Alex in the classroom in the Hogg Middle School Project kept students from becoming too frustrated with such glitches. In contrast, the understandable frustration teachers had with crashes in the 2005-2006 Project often spilled over to students; this was highly counterproductive.

Since the inception of the 2006-2007 Project in August 2006, the Reasoning Mind software system has been performing very reliably. The website availability to students and teachers has been in excess of 99%. There has been only one system breakdown; that lasted less than twenty minutes.

Rewards. The instructional system in the 2003 Project was built on the use of rewards. Students were told that "it's all about points" and the anticipation of rewards boosted student engagement. Certificates of recognition which were a very important part of the 2003 Project were not given during the 2005-2006 Project. During the Hogg Middle School Project rewards had a stimulating effect on student performance; they served as incentives. In the 2005-2006 Project, when used, rewards were more like an acknowledgement of past accomplishments. The timely use of rewards, the public acknowledgement of student success, and continuous encouragement provided to students during the Hogg Middle School Project created an atmosphere of excitement and competition (with himself or herself and with other students). Learning mathematics during the 2003 Project became doable, meaningful, and fun to students. A similar atmosphere was not created in most of the 2005-2006 Project classrooms. Many of the teachers commented on this. Clearly, when students became bored they were less likely to read the solutions, read the theory, and get the answers correct. This factor alone can make a perfectly sound curriculum useless.

The reward program has been given special attention in the 2006-2007 Project. The importance of earning points has been emphasized. Students are receiving postcards from the Genie acknowledging their academic achievement. Certificates of recognition used in the Hogg Middle School Project—and missing in the 2005-2006 Project—have been re-established. To create an atmosphere of excitement, teachers have been trained in using different strategies for running tournaments and competitions. Such competitions are effective in fighting boredom, elevating the level of student engagement, and focusing students on achieving the most important academic targets.

The Genie. The Genie was much more lively and relevant to students in the 2003 Hogg Middle School Project. The quality of e-mail responses to student letters was noticeably higher. The Genie challenged students by offering the option to solve harder problems, giving students the final decision of whether to go to a higher level of difficulty in problem solving or to stay at the same level. The Genie also acknowledged the achievement of higher levels of mastery by the students. In the 2005-2006 Project, the Genie usually just told students whether their problems were solved correctly or incorrectly.

The Genie's behavior in the system has not been made richer yet, due to the lack of required software development resources. This task has been postponed to the first half of 2007. However, efforts have been made to improve student-to-Genie communication by e-mail. Besides receiving responses to their questions e-mailed to the Genie, students are getting e-mail messages from the Genie acknowledging good progress and encouraging students to strive for higher performance.

Tutors. Tutors were available to help students in the 2003 Hogg Middle School Project whenever the students needed help. They were literally on stand-by, waiting for the system to suggest that a lesson must be given to a student. The student was not allowed to proceed until a tutor helped the student overcome a difficulty in understanding the material. However, clearly, there were not enough tutors in the 2005-2006 Project. Tutors were not available to clarify unclear solutions or walk the students through long solutions. This certainly hurt the project results.

The Reasoning Mind tutoring department has recruited and trained 25 volunteer and work-study tutors, each of whom is contributing from three to fifteen hours of tutoring per week. The department continues its recruitment efforts, hoping to hire an additional 25 tutors by the end of 2006. Staffed with about 50 tutors, the department will have enough resources to serve all needy students. This is a major improvement as compared to the lack of tutors experienced in the 2005-2006 Project.

Riddles and Games. Students in the Hogg Middle School Project had access to mathematics riddles and interactive games; these played a very important role in teaching the students non-standard thinking skills and they engaged the students instantly. No riddles or games were available to students in the 2005-2006 Project; consequently, their work in the system was rather monotonous, which often led to loss of interest.

For the 2006-2007 Project, a two-player game, MathRace, is fully operational and is being played by Reasoning Mind students. The game allows students to directly compete with each other in problem solving. In addition, the Riddle Machine is in the development and is scheduled for release by the Spring semester of 2007.

Supplemental Curriculum versus Core Curriculum. Although the Hogg Middle School Project was implemented as a supplement, Julia and Alex succeeded in having students develop a serious attitude toward their Reasoning Mind class. In contrast, most of the 2005-2006 classes were taught as traditional electives, where neither students nor teachers were expected to work hard and demonstrate noticeable achievement. Also, students were probably confused by two different approaches to mathematics given to them at the same time. The conceptually oriented approach of the Reasoning Mind curriculum clashed with the algorithmic approach of their regular mathematics classes; the different definitions and rules also confused them, since students did not know which to use.

During the 2006-2007 Project, most of the Reasoning Mind fifth grade student (about 200 out of 250) are taking the Reasoning Mind curriculum as the core curriculum. This significantly elevated the importance of the program in the eyes of both students and teachers. Also, taking the Reasoning Mind curriculum as a core prevents possible student confusion resulting from concurrent learning of mathematics in two very different ways--traditional and Reasoning Mind. About 250 sixth grade students are taking RM as a supplement. It is conceivable that although such an implementation is not optimal, students might still benefit from the program if it is implemented with fidelity.

Insufficient Time Which Resulted in a Lack of Review and Diagnostics. Many of the schools in the 2005-2006 Project did not start the Project until the spring semester. In addition, much of scheduled class time, sometimes up to 40%, was lost due to poor classroom management and school activities that took priority over Reasoning Mind classes. As a result, students were not given enough time to master the entire Reasoning Mind curriculum. In an effort to expose students to all objectives comprising the curriculum, the Project had to simplify the pedagogy used by the Genie by disabling reviewing of the previous topics and turning off the diagnostics of deficiencies in students' learning. Reviewing previous topics is important, since students forget easily. It is likely that the 2005-2006 Project would have been far more effective if the students had the opportunity to review topics on which they were weak, to master the skills that were prerequisites to the content presented later. Likewise, the absence of diagnostics meant that there was no re-teaching of those objectives that students failed to master. This hampered students' ability to understand material that required knowledge of previous objectives.

Homework. The homework module was not fully functional during the 2005-2006 Project. Homework is another way to add review. Also, homework is an opportunity to increase the amount of time students spends doing Reasoning Mind mathematics by as much as 50%, which could help partially remedy the shortage of time that students were facing.

During the 2006-2007 Project, fifth grade students who are taking Reasoning Mind as core have been given sixty to ninety minutes per day of instructional time. This time is sufficient for the most students to master the entire curriculum with fully enabled review and diagnostics. Students taking Reasoning Mind as a supplement are also enjoying the benefits of review and diagnostics. The homework module is currently fully functional. Teachers can assign individualized homework to students on a daily basis. The only limitation is that every homework assignment is limited to five problems.

Open Questions

In the view of Reasoning Mind staff, the following questions are the most important ones which remain open. That is, these are questions that should be answered in the course of the 2006-2007 Project and in future projects.

> To what extent can students with low reading skills be successful in mastering the Reasoning Mind mathematics curriculum?
>
> What is the minimum level of prerequisite mathematics competency required of students enrolled in the Reasoning Mind grade mathematics curriculum to ensure

mastery of the curriculum? What percent of students are prepared to take the Reasoning Mind curriculum?

Is there a significant difference in the level of mathematics competency required to meet state standards and the level of mathematics competency required to prepare students for the challenges of the high school mathematics curriculum?

To what extent will schools value giving their students deep mathematics knowledge if this knowledge does not directly translate into better performance on state standardized tests?

What is required from teachers to be effective in implementing the Reasoning Mind program? Can a sufficient number of such teachers be attracted to the Reasoning Mind program and trained appropriately?

What level of teacher support is required from Reasoning Mind to ensure the success of the project? Can this support be provided in a scalable way?

How effective is the teacher performance-based bonus program?

How effective is the student incentive program?

How scalable is the Reasoning Mind tutoring program? Will schools be able to provide the required tutoring resources?

How scalable is the Deployment Coordinator model? Can Deployment Coordinators be effectively trained to adequately supervise and support teachers?

References

Metiri Group. *Technology in Schools: What the Research Says.* San Jose, California: Cisco Systems, Inc., 2006.

Weber, W. A. “An Evaluation of the Reasoning Mind Pilot Program at Hogg Middle School.” Houston: University of Houston, 2003.