HOUSTON EDUCATION RESEARCH CONSORTIUM
KINDER INSTITUTE FOR URBAN RESEARCH
RICE UNIVERSITY

Analytic Summary of Existing *Reasoning Mind* Evaluations

UPDATED

March 16, 2012

SUMMARY

Reasoning Mind (RM) is a technology-based program designed to teach elementary and middle school students mathematics and logical reasoning skills. The program was piloted in one school in the Houston Independent School District (HISD) in 2003, and has since been adopted in selected schools within the district. This report is written for the purpose of providing HISD with information regarding the quality and strength of the existing evidence related to the effectiveness of RM in raising student achievement. This report uses the *What Works Clearinghouse* standards for research (2010) to evaluate the quality of the evidence generated to date on the effectiveness of RM in raising achievement. Our comprehensive analysis of each RM evaluation is listed in Table 1.

In the course of evaluating the existing evidence on RM, we found five studies that specifically addressed the effectiveness of RM in raising student achievement. Of these five studies, only one met evidence standards for research. The other four studies were compromised by two issues: 1) non-equivalence between experimental and control/matched groups, and 2) failure to adequately adjust for or account for non-equivalence in statistical models. Given that both problems occurred jointly in all four studies, we question the accuracy of the results presented in these four studies.

Based on the study that provided the most reliable evidence, HERC finds mixed evidence of the relationship between RM participation and mathematics achievement. Existing evidence suggests that RM participation does not significantly raise mathematics achievement for the group of students participating in the study, as measured by the mathematics section of TAKS, the state standardized achievement test. This same study found evidence for strong and positive effects related to RM and increased mathematics achievement, as measured by RM's own mathematics achievement test.[1]

METHOD

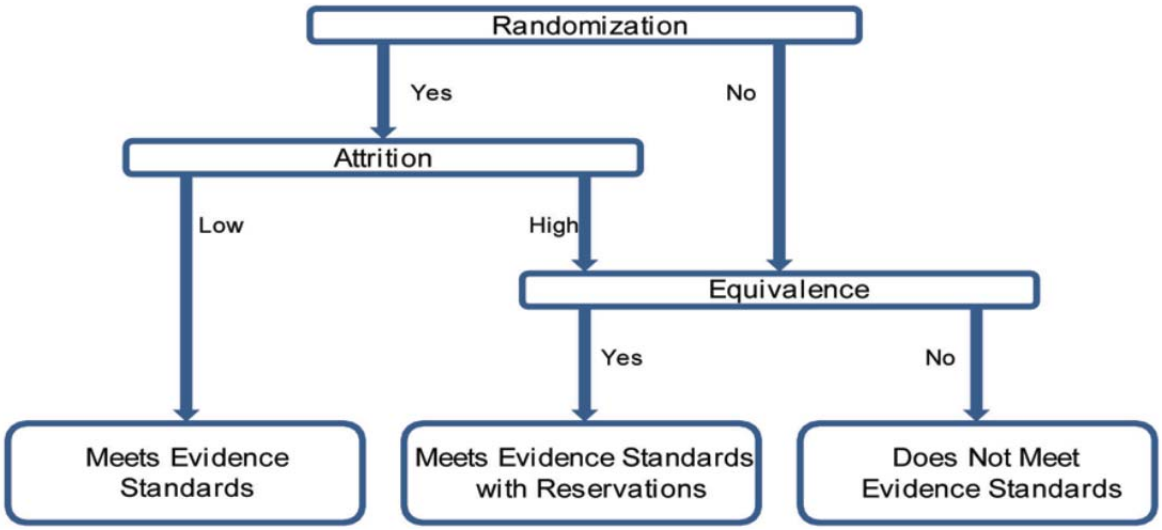
To conduct this evaluation, we identified all known existing studies of RM, including four independent evaluations that were conducted on behalf of RM and one HISD evaluation. Based on the *What Works Clearinghouse* guidelines, we identified each study as "Meets Evidence Standards", "Meets Evidence Standards with Reservations", or "Does Not Meet Evidence Standards". Criteria for each designation are demonstrated and described below.

Meets Evidence Standards: An evaluation that meets evidence standards includes randomization between treatment and control groups and low attrition.

Meets Evidence Standards with Reservations: An evaluation that meets evidence standards with reservations either includes randomization, but with high attrition, or does not include randomization but equivalence between treatment and comparison groups is adequately established.

[1] The RM mathematics achievement tests were developed in-house and designed specifically to assess mathematics achievement with regard to ratios, rates, and proportions. Even though the RM tests were not developed independently, they were independently evaluated and found to have sound psychometric properties (Weber, 2003).

Does Not Meet Evidence Standards: An evaluation that does not meet evidence standards lacks both randomization and equivalence. Furthermore, evaluations characterized by randomization and high attrition (particularly with high differential attrition) that fail to establish and adequately account for non-equivalence do not meet evidence standards.



As is stated by the *What Works Clearinghouse*, decisions about education programs, policies, and practices should be based on relevant studies that reach sound conclusions about the effectiveness of particular interventions. When studies about a particular intervention—such as RM—meet evidence standards (with or without reservations), educators can exercise confidence in conclusions about the effectiveness of the intervention.

Of the five existing studies that evaluate the influence of RM on student math achievement scores, we found one study (Waxman & Houston, 2008) that met the evidence standards (see Table 1). As a result, we recommend that HISD exercise confidence in the findings associated with this study. The other studies we evaluated did not meet evidence standards—either with or without reservations. In an effort to provide HISD with thorough information about each study, this report includes results from these studies; however, we suggest that caution must be exercised when interpreting these results. A detailed explanation of the weaknesses in each of these studies is listed in Appendix A.

The *What Works Clearinghouse* guidelines described above have been scrutinized for their strict equivalency guidelines for both experimental and quasi-experimental studies. We acknowledge these criticisms and are willing to exercise some leniency in our interpretation of evidence quality. In other words, we would be interested in reporting the evidence from high quality quasi-experimental studies that employ appropriate and rigorous methods—even if they fall short of the *What Works Clearinghouse* guidelines for equivalency. This, however, was not an issue in our summary of RM evaluations. While equivalence was an issue in all four of the studies that did not meet evidence standards, each of these studies utilized statistical methods that did not appropriately adjust for or account for differences between treatment and control/matched

groups. Accounting for such differences would likely change the outcomes reported in these studies.

THE EVIDENCE ABOUT REASONING MIND

Based on the one study that met the evidence standards (Waxman & Houston, 2008), RM is associated with the following achievement-based outcomes:

- No significant or meaningful gains in mathematics achievement as measured by the TAKS mathematics section.

- Significant and meaningful gains in mathematics achievement as measured by RMs own mathematics test.

Based on survey information from participating RM students and teachers, RM is also associated with several additional outcomes:

- In general, students liked RM and preferred learning math with RM rather than in a regular math class. Students also reported that they liked math more after participating in an RM class.

- Teachers reported high student engagement and interest in their RM classes. However, half of the teachers did not agree that RM benefited students with weak reading skills.

- Teachers also reported that RM was useful to them in improving their instruction and that it benefited students. Nevertheless, half of the teachers found it more difficult to teach an RM class than a traditional class.

- In spite of any difficulties associated with learning a new method of instruction, 100% of teachers agreed that they would like to continue using RM in their classroom.

While only one study incorporated a research design that would confidently yield credible results, the findings from the other studies were mixed in terms of RM raising achievement on TAKS. One small-scale, pilot study that did not meet evidence standards found that RM improves student performance on TAKS. Two other studies (neither of which met evidence standards) did not find any differences between RM and non-RM student mathematics achievement on TAKS. Yet another more recent study[2] (which also did not meet evidence standards) found small but significant increases in TAKS score, when comparing RM and non-RM students; however, these results disappeared when SES and racial/ethnic differences between the treatment and matched groups were accounted for.

[2] This study (Waxman & Houston, 2012) reported positive, significant differences between RM and non-RM students. This difference decreased (but was still statistically significant) when controlling for SES. However, further communication with the authors revealed that the increase in TAKS achievement scores associated with RM participation disappeared when racial/ethnic differences between the treatment and matched groups were accounted for through more thorough statistical analyses. Nevertheless, the study did not report the results of these analyses.

LESSONS FROM REASONING MIND EVALUATIONS

All educational programs require information about three components for their effectiveness to be fully realized:

Information about Recruitment: How are schools, classrooms, teachers, and students selected into the program?

Information about Retention: Who stays in the program, who drops out of the program, and why do some drop out?

Fidelity of Implementation: To what degree is the program appropriately used by teachers, and to what degree are students exposed to it?

Because information related to all three of these components was generally absent from all evaluations of RM, we do not have good information about *why* RM did not yield the anticipated results—specifically related to TAKS achievement. One study suggests that lack of significant differences between RM and non-RM students' mathematics achievement may be due to "lower content validity of TAKS to the Reasoning Mind curriculum that emphasizes problem solving and other higher-level thinking skills" (Waxman & Houston, 2008). The same study also suggests that TAKS scores may be characterized by lower stability and reliability from one year to the next, which may explain the non-significant results associated with TAKS. These TAKS-specific issues may or may not be less of an issue with the new STAAR assessments. Other studies highlighted the great variation across RM sites and classrooms in terms of recruitment, retention, and implementation (for example, see Weber, 2006). This variation was not accounted for in any of the existing RM evaluations.

While we strongly support the use of randomized research designs (as was used in many of the RM studies we reviewed), we suggest that such designs can be seriously compromised if researchers are unable to account for recruitment, retention, and implementation in both treatment and control groups. In such cases where randomization is not feasible, we recommend a high-quality quasi-experimental design where such differences can be accounted for. Most of the studies that did not meet evidence standards could have met standards with reservations, had they appropriately accounted for issues related to recruitment, retention, and implementation—particularly as these issues influence non-equivalency between treatment and control or matched groups.

RM RESPONSES TO RESEARCH

To better understand the research and conclusions of the numerous evaluations of RM, we met with RM to discuss their program, any changes the RM program has experienced over time, and their response to this summary report. Through our discussions with RM, we were informed about several contextual elements related to the implementation of RM that were not included in the research evaluations we reviewed.

- RM suggests that their program is aimed at improving students' fundamental knowledge of mathematics, problem-solving, and higher-order thinking skills—all of which are

preparatory for students' success in algebra. RM urges that TAKS does not sufficiently measure mastery of these fundamental principles of mathematics. However, RM suggests that the effects of participating in RM will become more salient when students enter higher grades and enroll in more advanced math classes (such as algebra). This has not yet been systematically tested; however, RM does have some preliminary support for this argument.[3] This may also highlight the potential disconnect between RM curriculum and types of mathematics knowledge measured on TAKS. To address this concern, RM developed their own mathematics achievement test, which is intended to offer an improved measure of students' mastery of fundamental principles of mathematics. RM highlights that—in the presence of proper implementation—most evaluations of their program demonstrated educationally meaningful gains in mathematics achievement relative to a control group when the RM test is used as the primary outcome of interest.

- During our conversations with RM, they consistently highlighted issues associated with relying on TAKS as an outcome. Even though each of the five studies of RM use TAKS as an outcome, RM suggests that TAKS is not the right instrument for measuring the effectiveness of the RM program. They argue that TAKS does not adequately measure the effectiveness of RM due to lower content validity, stability, and reliability of TAKS scores (Waxman & Houston, 2008). It is currently unclear whether or not these same issues will be associated with the STAAR assessment.

- RM has consistently hired third-party evaluators to analyze the successfulness of their program. They have also diligently used the outcomes and suggestions for improvement to make data-driven decisions about their program. For example, an early evaluation of RM demonstrated that uneven implementation severely limited the success of RM. Consequently, in the 2007-2008 school year, RM created an implementation department, hired a director for that department, and more fully developed a team of program coordinators specifically trained to support teachers as they implement RM in their day-to-day activities. Furthermore, these program coordinators assess the uniformity and quality of RM implementation. As of 2007, RM teachers also receive an intensive, 3-5 day training during the summer, as well as continued support through classroom observations and best practice workshops as they become familiar with RM. In summary, the implementation of RM has vastly improved over time, suggesting that the effectiveness of RM may improve as implementation improves. It is also important to note that three of the five RM evaluations (including the study that met evidence standards) pre-date these improvements in RM implementation.

[3] A preliminary internal RM report demonstrates that RM students in Angleton Independent School District perform better on the Orleans-Hanna Algebra Readiness Assessment than non-RM students. This difference was not tested for statistical significance.

CONCLUSIONS

RM is a mathematics program that is well-liked by both teachers and students. However, these positive attitudes and views of RM do not necessarily translate into consistent evidence regarding student achievement. Evidence reported in existing studies of RM suggests that RM does not significantly raise achievement on TAKS; however, RM is associated with significant and meaningful increases in achievement as is measured by the RM mathematics test. Nevertheless, these findings do not reflect RM's most recent changes in program implementation. It is also possible that RM participation improves students understanding of fundamental mathematic principles, which, in turn, better prepares students for more advanced mathematics such as algebra. However, this has not yet been tested systematically.

HERC Recommendations for Future Evaluation of Research

One conclusion HERC has drawn from this review process is that HISD has access to a great deal of research evaluating their programs, including both internal reports produced by HISD research staff and external reports produced by researchers outside of HISD. However, this research varies widely in quality, and thus in its ability to predict the effectiveness of the program. This variation is particularly consequential because these research reports are used to assist HISD leaders in making decisions regarding funding and program continuation. High-quality, unbiased research is most likely to provide an accurate representation of a program's effectiveness, uncover areas for improvement, and guide HISD leaders when setting budget priorities. In contrast, lower-quality research often leads to incorrect conclusions about the effectiveness of a program. As a consequence, all reports used for decision-making should be subject to a review process.

TABLE 1. SUMMARY OF REASONING MIND EVALUATIONS

Study	Source of Evaluation	Outcome Studied	Quality of Evidence[1]			Effectiveness			Applicability to HISD Population
			Overall Rating	Strength of Research Design	Analytic Method	Size of Impact & Statistical Significance	Effect Size & Substantive Significance	Effect on Achievement Gaps	
Waxman & Houston, 2008 ***"An Evaluation of the 2006-2007 Reasoning Mind Program"***	Independent evaluation conducted on behalf of Reasoning Mind.	RM Math Test	**Meets Evidence Standards**	**Meets Evidence Standards**	Appropriate, assuming equivalency of treatment and control groups.	Treatment group scored about 2-3 points higher than control group, varying slightly by school, a statistically significant finding.	Effect sizes ranged from 0.43 (a medium effect) to 0.80 standard deviations (a large effect). These effects are considered educationally meaningful.	Not examined.	Generalizable to Cornelius Elementary.
	Independent evaluation conducted on behalf of Reasoning Mind.	2007 TAKS Test (Math Section Only)	**Meets Evidence Standards**	**Meets Evidence Standards**	Appropriate, assuming equivalency of treatment and control groups.	No statistically significant differences between treatment and control groups.	No effect.	Not examined.	Generalizable to Cornelius Elementary.
Waxman & Houston, 2012 ***"Evaluation of the 2010-2011 RM Program in Beaumont ISD"***[2]	Independent evaluation conducted on behalf of Reasoning Mind.	2011 TAKS Test (Math Section Only)	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given non-equivalence problems, inappropriate for research design.	Treatment group scored about 12 points (38%) higher than matched group.[3]	Small, yet statistically significant effect. Not considered "educationally meaningful".	Not examined.	Generalizable to Beaumont ISD. BISD is smaller, but similar to HISD in racial/ethnic and SES composition.

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness			Applicability to HISD Population
			Overall Rating	Strength of Research Design	Analytic Method	Size of Impact & Statistical Significance	Effect Size & Substantive Significance	Effect on Achievement Gaps	
Waxman & Houston, 2012 ***"Evaluation of the 2010-2011 RM Program in Beaumont ISD"***	Independent evaluation conducted on behalf of Reasoning Mind.	2011 TAKS Test (Math Section Only) for RM students only	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Appropriate, but not a comparison of treatment and matched groups.	Percent of correct answers on RM curriculum and number of RM objectives completed were important predictors of math achievement.	Effect sizes (percentage of standard deviations) not reported; however, results were statistically significant.	Not examined.	Generalizable to Beaumont ISD. While BISD is much smaller than HISD, it's racial/ethnic and SES make-up is similar to HISD.
HISD, 2011 ***"Reasoning Mind Program Evaluation 2010-2011"***[4]	HISD Evaluation.	2011 TAKS Test Scores (both Spanish and English versions)	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given non-equivalence problems, inappropriate for research design.	No statistically significant differences between treatment and control groups.	No effect.	Not examined.	Inferred that results may be generalizable to students in RM schools in HISD.
	HISD Evaluation.	2011 TAKS Passing Standards	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Descriptive comparisons of treatment and matched groups only. Inappropriate for research design.	N/A	N/A	Economic advantage, LEP, and race gaps examined for treatment group only.	Inferred that results may be generalizable to students in RM schools in HISD.

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness			Applicability to HISD Population
			Overall Rating	Strength of Research Design	Analytic Method	Size of Impact & Statistical Significance	Effect Size & Substantive Significance	Effect on Achievement Gaps	
HISD, 2011 *"Reasoning Mind Program Evaluation 2010-2011"*	HISD Evaluation.	2011 TAKS Commended Standards	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Descriptive comparisons of treatment and matched groups only. Inappropriate for research design.	N/A	N/A	Economic advantage, LEP, and race gaps examined for treatment group only.	Inferred that results may be generalizable to students in RM schools in HISD.
	HISD Evaluation.	Stanford 10 Test	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given non-equivalence problems, inappropriate for research design.	Results are inconsistent. Not enough information about statistical tests presented in the tables to resolve inconsistencies.	Unresolvable inconsisten-cies in analyses. Cannot determine effect sizes.	Not examined.	Inferred that results may be generalizable to students in RM schools in HISD.

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness			Applicability to HISD Population
			Overall Rating	Strength of Research Design	Analytic Method	Size of Impact & Statistical Significance	Effect Size & Substantive Significance	Effect on Achievement Gaps	
Weber, 2006 *"An Evaluation of the 2005-2006 Reasoning Mind Project"*	Independent evaluation conducted on behalf of Reasoning Mind.	2006 TAKS Test (Math Section Only)	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given vast implement-ation and evaluation problems highlighted in the study, inappropriate for research design.	Results vary by school, grade, and quality of research design. General finding is one of no consistent, statistically significant effect.	General finding of no effect.	Not examined.	Inferred that results may be generalizable to students in RM schools in HISD.
	Independent evaluation conducted on behalf of Reasoning Mind.	Stanford 10 Test (Math Section Only	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given vast implement-ation and evaluation problems highlighted in the study, inappropriate for research design.	Results vary by school, grade, and quality of research design. General finding is one of no consistent statistically significant effect.	General finding of no effect.	Not examined.	Inferred that results may be generalizable to students in RM schools in HISD.
Weber, 2003 *"An Evaluation of the RM Pilot Program and Hogg Middle School"*[5]	Independent evaluation conducted on behalf of Reasoning Mind.	RM Math Test	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given attrition problems, inappropriate for research design.	Treatment group scored 5.66 points higher than control group, a statistically significant finding.	Effect Size = 1.66 standard deviations, which is very large and considered educationally meaningful.	Not examined.	Not generalizable to HISD population.

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness			Applicability to HISD Population
			Overall Rating	Strength of Research Design	Analytic Method	Size of Impact & Statistical Significance	Effect Size & Substantive Significance	Effect on Achievement Gaps	
Weber, 2003 ***"An Evaluation of the RM Pilot Program and Hogg Middle School"***	Independent evaluation conducted on behalf of Reasoning Mind.	2003 TAKS Test (Math Section Only)	**Does Not Meet Evidence Standards**	**Does Not Meet Evidence Standards**	Given attrition problems, inappropriate for research design.	Treatment group scored 5.1 points than control group, a statistically significant finding.	Effect Size = 0.79 standard deviations, which is large and considered educationally meaningful.	Not examined.	Not generalizable to HISD population.

[1] Quality of Evidence evaluations based on the *What Works Clearinghouse Procedures and Standards Handbook (Version 2.1)* (see pages 11-18).
[2] Study did not qualify for a "Meets Standards with Reservations" designation due to the absence of discussion regarding equivalence between treatment and control groups and also due to an absence of accounting for observable differences in treatment and control groups in the analyses.
[3] While this study found statistically significant improvements in TAKS scores related to RM participation, it is important to note that the treatment and "control" groups were not well matched in terms of student background characteristics. While they authors did account for student SES in their statistical models, they did not account for race/ethnicity or other observable characteristics that may influence the outcomes of their analyses. Communication with the authors revealed that the increase in TAKS scores attributed to RM participation disappeared when racial/ethnic differences between the treatment and matched groups were accounted for. This suggests that the supposed increased TAKS scores attributed to RM participation is likely related to the imbalance of student background characteristics between the treatment and matched groups. As a result, the outcomes reported in this study should be interpreted with caution.
[4] With a few additional statistical adjustments and procedures, this study could easily be rated as "Meeting Standards with Reservations." HERC is willing to offer more detailed comments and assistance, should HISD wish to improve this study.
[5] Information about this study supplemented with information provided in the "*Description and Evaluation of Reasoning Mind's 2003 Pilot Project*" (September 2003), an internal report generated by Reasoning Mind.

APPENDIX A: METHODOLOGICAL SUMMARY OF REASONING MIND EVALUATIONS THAT DO NOT MEET EVIDENCE STANDARDS

Study

Waxman & Houston, 2012 *"Evaluation of the 2010-2011 Reasoning Mind Program in Beaumont ISD"*

Outcome

2011 TAKS Test (Math Section Only)

Randomly Assigned

No.

Equivalence

Baseline Data: Yes, pretest achievement data.

Adequate Control Variables: No. While controls for pretest and SES differences were included, other observable differences were not accounted for in analyses.

Appropriate Matching: No. Non-equivalence between treatment and matched sample that is not accounted for in analyses.

Attrition

Overall Attrition: Not specified. Not assumed to be a problem.

Evenness between groups: Not specified. Inferred that students in treatment group were generally more advantaged than matched group.

Generalizability

Treated sample size & generalizable population: Not specified. Inferred to be generalizable to students in RM schools.

How representative of HISD population: Not specified; however, Beaumont ISD has demographic characteristics that are similar to HISD.

Analytic Method

Not appropriate. Non-equivalent treatment and matched groups that are not accounted for in analyses.

Study

HISD, 2011 *"Reasoning Mind Program Evaluation 2010-2011"*

Outcome

2011 TAKS, TAKS Passing and Commended Standards, & Stanford 10 Test Scores

Randomly Assigned

No. Matched sample.

Equivalence

Baseline Data: Yes, pretest achievement data.

Adequate Control Variables: No. No control variables used to adjust for non-equivalence between treatment and matched samples.

Appropriate Matching: No. Non-equivalence between treatment and matched samples (particularly for LEP students).

Attrition

Overall Attrition: Not specified. Not assumed to be a problem.

Evenness between groups: N/A

Generalizability

Treated sample size & generalizable population: Not specified. Inferred to be generalizable to students in RM schools.

How representative of HISD population: Not specified. Inferred to provide some generalizability to underperforming students in low-income, high-minority schools who remain in the same school for the duration of the school year.

Analytic Method

Not appropriate. Non-equivalent treatment and matched groups. Inappropriate use of statistical tests. Lack of descriptive comparison between treatment and matched groups.

Study

Weber, 2006 *"An Evaluation of the 2005-2006 Reasoning Mind Project"*

Outcome

2006 TAKS Test (Math Section Only) & Stanford 10 Test (Math Section Only)

Randomly Assigned

In many cases yes, but compromised by non-random selection, non-equivalent control groups, and lack of school administrative cooperation.

Equivalence

Baseline Data: Yes, pretest achievement data. Missing data on pretest for many students due to high student mobility and lack of school administrative cooperation.

Adequate Control Variables: No. While typically unnecessary with random assignment, the limitations of this study suggest a need for control variables to ensure equivalence.

Appropriate Matching: No. While typically unnecessary with random assignment, the limitations of this study suggest a need for information related to the quality of match between control and treatment groups.

Attrition

Overall Attrition: Not specified. Inferred to be high.

Evenness between groups: Not specified. Inferred to be higher in control groups than in treatment groups.

Generalizability

Treated sample size & generalizable population: Not specified. Inferred to be generalizable to students in RM schools.

How representative of HISD population: Not specified. Inferred to provide some generalizability to underperforming students in low-income, high-minority schools who remain in the same school for the duration of the school year.

Analytic Method

Not appropriate based on lack of equivalence in treatment and control groups. Treatment and control groups generally too small to draw meaningful conclusions from analyses (less than 30 students).

Study

Weber, 2003 *"An Evaluation of the Reasoning Mind Pilot Program and Hogg Middle School"*

Outcome

RM Math Test & 2003 TAKS (Math Sections Only)

Randomly Assigned

Yes, but compromised by scope (62 students total), by size of treatment and control groups (30 and 26 students, respectively, and by differential attrition.

Equivalence

Baseline Data: Yes, pretest achievement data.

Adequate Control Variables: No. While typically unnecessary with random assignment, the limitations of this study suggest a need for control variables to ensure equivalence.

Appropriate Matching: No. While typically unnecessary with random assignment, the limitations and differential attrition in this study suggest a need for information related to the quality of match between control and treatment groups.

Attrition

Overall Attrition: 10% (Acceptable).

Evenness between groups: 20% in Control Group (Very High). 0% in treatment group. After attrition, control group too small for accurate comparison (less than 30 cases).

Generalizability

Treated sample size & generalizable population: 30 Students; Generalizable to 62 "C" students who did well enough on RM pre-selection test to qualify for RM intervention.

How representative of HISD population: Provides poor representation of HISD's population.

Analytic Method

Not appropriate with high attrition in control group. No demonstration of equivalence between treatment and control groups. Control group too small to draw meaningful conclusions from analysis after attrition.

REFERENCES

Houston Independent School District. (2011). *Reasoning Mind Program Evaluation 2010-2011*. Houston, TX.

Waxman, H. C. & Houston, W. R. (2012). *Evaluation of the 2010-2011 Reasoning Mind Program in Beaumont ISD*. Reasoning Mind, Inc., Houston, TX.

Waxman, H. C. & Houston, W. R. (2008). *An Evaluation of the 2006-2007 Reasoning Mind Program*. Reasoning Mind, Inc., Houston, TX.

Reasoning Mind. (2003). *Description and evaluation of Reasoning Mind's 2003 Pilot Project*. Reasoning Mind, Inc. Houston, TX.

What Works Clearinghouse. (2011). *What Works Clearinghouse Procedures and Standards Handbook*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

Weber, W. A. (2003). *An Evaluation of the Reasoning Mind Pilot Program at Hogg Middle School*. University of Houston, Houston, TX.

Weber, W. A. (2006). *An Evaluation of the 2005-2006 Reasoning Mind Project*. Reasoning Mind, Inc., Houston, TX.