The Effects of Solve It!
on Seventh-Grade Students’ Math Problem Solving
Year 3 Results

This research is supported by grant No. R324A070206 from the Institute of Education Sciences (IES)
U.S. Department of Education.

Marjorie Montague, Ph.D.
University of Miami
mmontague@miami.edu

Craig Enders, Ph.D.
Arizona State University

Samantha Dietz, Ph.D.
University of Miami

www.education.miami.edu/solveit
Purpose:
To improve mathematical problem solving for middle school students by providing general education math teachers with a research-based instructional program.

Sample:
Year 3: Efficacy Study: 20 pairs of middle schools matched on FCAT performance grade and SES. One school from each pair randomly assigned to intervention.

Intervention:
*Solve It!* implemented for seven months with periodic progress monitoring.
3 Primary Research Questions

1. What are the effects of *Solve It!* on math problem solving as measured by curriculum-based measures of problem solving (CBM) for the intervention and comparison group students and subsets of students (i.e., students with learning disabilities - SLD, low achieving students – LAS, and average-achieving student – AAS)?

2. What are the effects of *Solve It!* on math problem-solving self-efficacy as measured by the Math Problem-Solving Self-Efficacy Scale (MPS-SES) for the intervention and comparison group students and subsets of students (i.e., SLD, LAS, AAS)?
3. What are the effects of *Solve It!* on math performance as measured by the Florida Comprehensive Assessment Test (FCAT) math and reading tests for the intervention and comparison group students and subsets of students (i.e., SLD, LAS, AAS)?
**Solve It! – A Routine for Solving Math Problems**

**Cognitive Processes/Self-Regulation Strategies**

**READ** (for understanding)

*Say:* Read the problem. If I don’t understand, read it again.

*Ask:* Have I read and understood the problem?

*Check:* For understanding as I solve the problem.

**PARAPHRASE** (your own words)

*Say:* Underline the important information. Put the problem in my own words.

*Ask:* Have I underlined the important information? What is the question? What am I looking for?

*Check:* That the information goes with the question.

**VISUALIZE** (a picture or a diagram)

*Say:* Make a drawing or a diagram. Show the relationships among the problem parts.

*Ask:* Does the picture fit the problem? Did I show the relationships?

*Check:* The picture against the problem information.

**HYPOTHESIZE** (a plan to solve the problem)

*Say:* Decide how many steps and operations are needed. Write the operation symbols (+, -, x, and /).

*Ask:* If I …, what will I get? If I …, then what do I need to do next? How many steps are needed?

*Check:* That the plan makes sense.
### Math Problem Solving Processes and Strategies

<table>
<thead>
<tr>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESTIMATE</strong></td>
<td><strong>predict the answer</strong>&lt;br&gt;Say: Round the numbers, do the problem in my head, and write the estimate.&lt;br&gt;Ask: Did I round up or down? Did I write the estimate?&lt;br&gt;Check: That I used the important information.</td>
</tr>
<tr>
<td><strong>COMPUTE</strong></td>
<td><strong>do the arithmetic</strong>&lt;br&gt;Say: Do the operations in the right order.&lt;br&gt;Ask: How does my answer compare with my estimate? Does my answer make sense? Are the decimals or money signs in the right places?&lt;br&gt;Check: That all the operations were done in the right order.</td>
</tr>
<tr>
<td><strong>CHECK</strong></td>
<td><strong>make sure everything is right</strong>&lt;br&gt;Say: Check the plan to make sure it is right. Check the computation.&lt;br&gt;Ask: Have I checked every step? Have I checked the computation? Is my answer right?&lt;br&gt;Check: That everything is right. If not, go back. Then ask for help if I need it.</td>
</tr>
</tbody>
</table>

*From Montague (2003). Copyright by Exceptional Innovations. Permission to photocopy this figure is granted for personal use only.*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Intervention (n=644)</th>
<th>Comparison (n=415)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLD</td>
<td>58 (9)</td>
<td>28 (7)</td>
</tr>
<tr>
<td>LAS</td>
<td>417 (65)</td>
<td>293 (70)</td>
</tr>
<tr>
<td>AAS</td>
<td>169 (26)</td>
<td>94 (23)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>266 (41)</td>
<td>187 (45)</td>
</tr>
<tr>
<td>Females</td>
<td>378 (59)</td>
<td>228 (55)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>38 (6)</td>
<td>17 (4)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>405 (63)</td>
<td>278 (67)</td>
</tr>
<tr>
<td>Black</td>
<td>201 (31)</td>
<td>120 (29)</td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>501 (77)</td>
<td>345 (83)</td>
</tr>
<tr>
<td>No</td>
<td>143 (23)</td>
<td>70 (17)</td>
</tr>
</tbody>
</table>
Curriculum-based Measures (CBM)

- Set of seven tests of one-, two-, and three-step math word problems.
  - Minimal if any prior specific mathematical concepts/knowledge required.
  - Whole numbers and decimals and four basic operations.
- Equated scores using the dichotomous Rasch model. Longitudinal analyses of CBM are appropriate only when test scores from repeated administrations have been vertically equated to a common scale score (Montague, M., Penfield, R., Enders, C., & Huang, J. (2010). Curriculum-based measurement of math problem solving: A methodology and rationale for establishing equivalence of scores. *Journal of School Psychology, 48*, 39-52. [http://dx.doi.org/10.1016/j.jsp.2009.08.002]).
- Administered seven measures (Oct – June).
  - Reliability = .71, .74, .76, .73, .75, .74, .79
Three-step problem

A store sells shirts for $13.50 each. On Saturday, it sold 93 shirts. This was 26 more than it had sold on Friday. How much did the store charge for all the shirts sold on both days?
Solve-It Practice Questions

Standard: MA.A.1.3.2

- Compares and orders integers, fractions, and decimals, numbers with exponents and numbers expressed as percents or in scientific notation including ordering on a number line.

Problem Source: Prentice Hall Mathematics (Course 2)
Page 250, TG: Assess: Problem 2

Problem

A 20-lb bag of dog food costs $21.50. A 30-lb bag of the same dog food costs $32.90. Find each unit price. Then determine the better buy.

Schematic Representation

<table>
<thead>
<tr>
<th>Price</th>
<th>$32.90</th>
<th>× $/lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>30 lb</td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>$21.50</td>
<td>× $/lb</td>
</tr>
<tr>
<td>Size</td>
<td>20 lb</td>
<td></td>
</tr>
</tbody>
</table>

30 lb - $32.90  20 lb - $21.50

Answer: $1.08 per pound; $1.10 per pound
The 20 pound bag is the better buy.
Math Problem-Solving Self-Efficacy Scale (MPSSSES)

- Modeled after the Self-Efficacy for Learning Scale (SELS; Zimmerman, Bandura, and Martinez-Pons, 1992).
- Measures academic self-efficacy beliefs specific to the domain of math problem solving.
- Original 10 items were subjected to a factor analysis and then reduced to 6 items that reflect one of Bandura’s sources of self-efficacy, “mastery experience.”
- Items measure perceived capability to perform well in math problem solving by endorsing items on a 4-point Likert scale.
- For the current IES study, alpha values were good at .85 (Wave 1), .86 (Wave 3), .90 (Wave 5), and .91 (Wave 7).
MPSES

1) I am certain that I can solve most math word problems.
   A. Not at all true  B. Hardly true  C. Mostly true  D. Absolutely true

2) I am confident that I can deal with solving most math word problems.

3) Thanks to my knowledge and skills, I can handle most math word problems.

4) I can remain calm when solving math word problems because I can rely on my knowledge and skills.

5) When I am confronted with a math word problem, I can usually find the right solution.

6) I can solve most math word problems.
Florida Comprehensive Assessment Test

- March 2009 and March 2010 scores
Procedures

INTERVENTION GROUP

- Professional development (3 days in August)
- Pretests and baseline CBM (CBM 1)
- District curriculum with embedded Solve It! Instruction (scripted lessons)

Treatment Fidelity

3 days of intensive instruction (average 97%, range 90% - 100%; agreement average 99%, range 94% - 100%)

Weekly practice sessions

(average 93%, range 77% - 100%; agreement average 99%, range 88% - 100%)

- Six additional CBM (monthly)
- Posttests
COMPARISON GROUP

- Pretests and baseline CBM (CBM 1)
- District curriculum following district pacing guide
  Teachers were asked to focus on problem solving one day per week.
  Observations (one in the fall and one in the spring)
- Three additional CBM (bi-monthly)
- Posttests
ANALYSES AND RESULTS
The data were consistent with a 3-level model where repeated measures (level 1) were nested within students (level 2) and students were nested within schools (level 3).

A linear growth model was the better fit for the analyses. We controlled for student-level covariates (ability level, ethnicity, gender, SES, FCAT math and reading scores) and school-level covariates (% free/reduced lunch, % minority, average FCAT math and reading scores)*.

*The model without the covariates produced identical substantive conclusions and nearly identical growth rate estimates.

For the final analyses (i.e., FCAT), the model was not estimable because we examined only two years of standardized achievement data (estimating the full set of variances and covariances requires at least three waves of data). Consequently, we simplified the model by removing the slope residuals.
CBM Results

- Students receiving *Solve It!* instruction \((n = 644)\) made *significantly greater growth in math problem solving* over the school year than students in the comparison group \((n = 415)\) as measured by curriculum-based measures of textbook-type problems. (medium effect size - Cohen’s \(d = .502\)).
Figure C1. Model-implied growth trajectories (averages) from the final CBM model. The growth curves are simple slopes that represent the expected developmental trajectory for individuals scoring at the covariate means.
Next we explored whether the effect of the intervention differed across (i.e., was moderated by) ability groups (i.e., SLD, LAS, AAS) – same covariates.

- Although scores differed (SLD < AAS by about 2.5 points – small effect size, $d = .267$), the intervention had a uniform impact across ability levels on the CBM.

- That is, ability group had no material impact on baseline differences, developmental rates, or the magnitude of the intervention effect.
Students receiving *Solve It!* instruction \((n = 644)\) made *significantly greater growth in math problem-solving self-efficacy* over the school year than students in the comparison group \((n = 415)\) as measured by the MPSSSES. (small effect size for rate of growth for the intervention group - Cohen’s \(d = .376\)).
Figure C2. Model-implied growth trajectories (averages) from the final self-efficacy model. The growth curves are simple slopes that represent the expected developmental trajectory for individuals scoring at the covariate means.
Next we explored whether the effect of the intervention differed across (i.e., was moderated by) ability groups (i.e., SLD, LAS, AAS) – same covariates.

- Although scores differed (SLD < AAS by about 1 point – small effect size, $d = .391$), the intervention had a uniform impact across ability levels on the MPSSES.
- That is, ability group had no material impact on baseline differences, developmental rates, or the magnitude of the intervention effect.
FCAT Results

- Students receiving *Solve It!* instruction \((n = 644)\) made *significantly greater growth on the FCAT math test between 2009 and 2010* than students in the comparison group \((n = 415)\) (small effect size for rate of growth for the intervention group - Cohen’s \(d = .216\)).

- Whereas, both groups improved on the FCAT *reading* test between 2009 and 2010 at about the same rate (small effect sizes for rate of growth, i.e., intervention, Cohen’s \(d = .334\); comparison, Cohen’s \(d = .256\)).
Next we explored whether ability group (i.e., SLD, LAS, AAS) moderated the effects of Solve It! on FCAT math scores.

- Scores differed (SLD < AAS by about 52 points and LAS < AAS by about 44 points – large effect sizes, $d = 1.68$ and 1.41, respectively).
- Analyses suggested that the three ability groups showed differences in their overall FCAT math averages and in their change rates, but ability had no material impact on the magnitude of the intervention effect. That is, the intervention influence was uniform across the three ability groups.
Figure C3. Model-implied growth trajectories (averages) from the FCAT mathmodel. The growth curves are simple slopes that represent the expected developmental trajectory for individuals scoring at the covariate means.
Thank you!