CURRICULUM & TEACHING STUDIES | RESEARCH ARTICLE

Assessing a multi-component math intervention within a cognitive-behavioral framework on the word problem-solving responses of a diverse group of third graders

Sheri Kingsdorf1* and Jennifer Krawec2

Abstract: In third grade the focus on math word problems becomes prominent. In the limited third grade research, teacher-mediated explicit instruction with multiple exemplars, teaching students to use visual representations, and the incorporation of self-strategies, have proven effective. For these practices to reach their full potential though, their content must be relevant and provide for growth to mature mathematical concepts. Based on these conclusions, additional research was needed. Therefore, the focus of this study was to evaluate the effectiveness of a multi-component word problem-solving intervention with explicit instruction strategies, multiple exemplars, the teaching of student-generated visual representations, incorporation of a self-monitoring checklist, and Common Core State Standards’ appropriate curriculum. Within a multiple baseline across behaviors design, the study evaluated the paraphrasing, visualizing, and computing word problem-solving responses of 10 third graders identified as learning disabilities, at-risk, and/or ESOL. The study revealed that all students made gains in some behaviors related to problem solving. Results are discussed in relation to a cognitive-behavioral framework and individual student characteristics, including discussions of limitations and educational significance.

ABOUT THE AUTHOR
Sheri Kingsdorf runs an applied behavior analysis consultation company that provides behavioral and educational support to staff working in public and private schools, home-based therapy settings, and day-treatment centers. She remains active in academia through adjunct teaching positions and collaborative research projects with colleagues. Her research interests include behavior analytic intervention research related to skill acquisition in the populations of children with autism, learning disabilities, and English Language Learners (ELL). She hopes to further investigate the cognitive-behavioral framework introduced in this study within other academic subjects; additionally studying the role that Skinner’s Verbal Behavior plays in the cognitive and behavioral processes associated with problem solving and comprehension.

PUBLIC INTEREST STATEMENT
Even for some adults, solving math word problems can be daunting. This history of struggle with word problems often begins in third grade, when they are first introduced into the curriculum. Unfortunately, the intimidation that surrounds this tough academic task can result in even teachers finding themselves without the necessary skills to address student learning in this area. Therefore, the aim of this study was to investigate a model of teaching math word problems to third graders that could be easily utilized by educators in the inclusive classroom, measure areas of learning beyond mathematical computation, and provide strategies which can continue to be used as word problems increase in complexity. The intervention outlined specific teaching procedures, taught students how to use paraphrasing and visualizing when approaching the problems, and encouraged students to monitor their solving process. Overall, we found that the students and the teacher responded favorably to the program.
1. Theoretical framework

Solving word problems is a complicated and often misunderstood process (Lai, Griffin, Mak, Wu, & Dulhunty, 2001). Over the years, prominent cognitive psychologists (e.g. Mayer, 1983; Pólya, 1957) have proposed organizational structures with embedded cognitive processes to aid in defining this ambiguous area. Synthesizing these structures and frameworks, the Organization for Economic Co-operation and Development (OCED) has outlined the following problem-solving processes: (1) understanding the problem, (2) characterizing the problem, (3) representing the problem, (4) solving the problem, (5) reflecting on the solution, and (6) communicating the problem solution (OCED, 2003). Figure 1 depicts the components of the OCED problem-solving model, including the processes, their variables, and their hypothesized relationships. Students must learn to navigate these processes in order to be successful in math problem solving, and thus teachers must find means of assessing students’ proficiency in each component. To do so, the components of the problem-solving process must be translated into observable terms. The behavior analytic perspective on problem solving helps to create this translation.

In the seminal work of Gagne (1959), problem solving is described as occurring in five phases: (1) reception of the stimulus situation, (2) concept formation of which rules to adopt to apply to parts of the stimulus situation, (3) planning a course of action, (4) decision-making within the course of action, and (5) verification of process, by receiving feedback that the solution was correct or incorrect. There is clear overlap between these cognitive and behavioral models of problem solving. When merged, as depicted in Figure 2, a more comprehensive and measurable view of problem solving emerges. The process in the cognitive-behavioral model, as in the cognitive model, is depicted in a linear fashion. For novice problem-solvers, according to the cognitive-behavioral model, a student attempting to solve a word problem would move through the process in the following manner:

- Reading the word problem (reception);
- Identifying the important information and constructing a problem representation by generalizing previously learned rules for solving problems, and discriminating differences in the target problem from previously targeted problems (concept formation);
- Making a plan to solve by choosing an operation/operations and identifying the number of steps needed to solve the problem (planning); identifying operations and steps is also rule-governed and closely related to the concept formation process;
• Checking the plan with the problem information to make sure that the plan matches the problem information and the representation (reflection). This step cohesively brings the concept formation and planning phases together;
• Showing the work on the paper, writing down the answer (communication); and
• Checking the work and the solution by using a calculator, reviewing the steps, or using inverse operations (verification).

However, a more experienced problem-solver may move through the process in a non-linear fashion. For example, he or she may be able to skip the phase of drawing a representation of the problem, moving directly from reception to planning.

Within this cognitive-behavioral problem-solving framework, a number of behaviors need to be observed for educators to assess the problem-solving skills of students. Determining if students understand how to solve the problem is made observable by evaluating their concept formation and planning behaviors (i.e. the information they identify as relevant, the representation that they construct, operation(s) they select, and steps identified as needed). This complex combination of skills makes problem solving a difficult skill to learn and an even more difficult skill to teach.

2. Introduction and review of the literature

Sets of standards drive the curricular content focus on numbers and operations, algebra, geometry, measurement, data analysis, and probability, as well as the process focus on problem solving, connections, communication, reasoning and proof, and representation (NCTM, n.d.). In addition, the introduction of Common Core State Standards (CCSS) has strengthened this focus (Common Core State Standards Initiative, 2012). Expectations have risen to include the application of problem-solving skills to real-life scenarios. These take the form of word problems, and are present as early as kindergarten (Common Core State Standards Initiative, 2012). Although word problems are introduced early, their weight in the curriculum increases in third grade, when high-stakes testing often begins. This is mismatched however, as many third grade textbooks do not adequately address problem solving (Jitendra et al., 2005). Additionally, the majority of problem-solving research is focused on the upper grades (e.g. Montague, Enders, & Dietz, 2011; Schoeffer Whity, 2009; Xin, Jitendra, & Deatline-Buchman, 2005). With student proficiency in problem solving expected at the third grade level, finding ways to promote student success in this area is imperative.

The body of research that exists on math problem solving in the upper grades has provided valuable insight into the difficulties students experience when attempting to solve word problems, as well as the intervention strategies that address them. Teachers have identified word problems as the most prominent area of student difficulty; this is especially the case for students identified with learning disabilities (LD) (Bryant & Bryant, 2008). Montague and Applegate (1993) conducted a study to investigate the problem-solving characteristics of students with LD. They found that
problem-solving errors were more related to struggles with representing the problem and planning 
a solution, as opposed to operation errors. Additional research has found that students with LD 
struggle with word problems as they increase in complexity, making more errors and demonstrating 
less productive solving practices (Kingsdorf & Krawec, 2014; Rosenzweig, Krawec, & Montague, 
2011). Problem complexity increases with the addition of steps, operations, and irrelevant informa-
tion in the problems, translating into issues of problem type and missing information (Powell, Fuchs, 
Fuchs, Cirino, & Fletcher, 2008). Unsurprisingly, middle-school students who are also English 
Language Learners (ELL) struggle with the linguistic complexity of word problems (Barbu & Beal, 
2010). This cross-curricular nature of word problems is a challenge for ELL students who may be 
struggling with reading comprehension. Overall, word problems are mathematically and linguisti-
cally complicated, increase in complexity throughout grade levels, and pose a challenge to many 
students.

As a result, numerous interventions have worked to remediate these areas of student struggle. 
Assistive technology, including the use of computers, audio-video devices, and calculators, has ef-
effectively increased the arithmetic and algebra word problem-solving skills of middle-school students 
with varying abilities (e.g. Bottge, 1999; Bottge & Hasselbring, 1993; Bottge, Heinrichs, Chan, Mehta, 
& Watson, 2003; Bottge, Rueda, Serlin, Hung, & Kwon, 2007). Strategy-training instruction, based on 
explicit teaching, using metacognitive strategies, and/or mnemonic devices, is another intervention 
that has successfully increased the word problem-solving skills of diverse groups of adolescents (e.g. 
Coughlin & Montague, 2011; Krawec, Huang, Montague, Kressler, & Melia de Alba, 2013; Maccini & 
Hughes, 2000; Maccini & Ruhl, 2000; Montague et al., 2011). Teaching the use of self-strategies has 
also proved a beneficial intervention component for increasing the correct problem-solving behavior 
of middle-school students (e.g. Case, Harris, & Graham, 1992; Montague, 2008). Problem structure or 
visual representation teaching procedures, targeting explicitly teaching students to represent the 
problem with a diagram or mathematical model, is yet another established intervention practice for 
bettering the word problem-solving skills of middle-school students across ability groups (e.g. 
Jitendra, DiPipi, & Perron-Jones, 2002; Jitendra, Hoff, & Beck, 1999; Jitendra et al., 2009; Xin et al., 
2005). Additionally, when looking at word problem-solving strategies including middle-school stu-
dents, recent meta-analyses support the use of visual representation teaching procedures (Xin & 

While some of the intervention strategies applied in the upper grades can be applied in the lower 
grades, exact replication is not always possible. Developmentally, third graders are different from 
middle-school students, where the majority of research has been targeted. From the perspective of 
cognitive development, as related to Piagetian theory (e.g. Piaget, 1970) third graders are beginning to 
develop concrete operations related to seriation and classification, but require concrete mathe-
matical activities with various representations to make connections and move closer to abstract 
thought. In contrast, middle-school students, moving into early adolescence, are beginning to de-
velop abstract reasoning, honing their ability to clarify important information needed in the problem, 
making hypotheses and inferences in problem solving, and evaluating their mathematical work 
(Ojose, 2008). These differences are apparent when considering an area like visual representation 
instruction in solving word problems. As such, third graders may benefit more from interventions 
focusing on using concrete objects to make visual representations (e.g. drawing a picture where 
each piece of important information is explicitly represented). Alternatively, middle-school students 
may be comfortable with the use of more abstract problem representations (e.g. drawing a diagram 
that uses a number sentence).

In addition to the developmental differences, curricular goals differ from grade to grade. At the 
third grade level, the Common Core State Standards calls for students to solve problems using all 
four operations (addition, subtraction, multiplication, and division), fractions, measurement, estima-
tion of time, liquid volume, masses of objects, and geometric measurement (Common Core State 
Standards Initiative, 2012). These skills are vastly different from the Common Core State Standards’ 
targets at the middle-school level, where students are expected to be learning proportional
relationships, expressions and linear equations, solving problems using area and volume, equations, using functions to describe quantitative relationships, and applying advanced geometric theories (Common Core State Standards Initiative, 2012).

Specific research teams that are targeting third graders (e.g. Fuchs and colleagues and Jitendra and colleagues) have addressed some of these limitations in the problem-solving research. The work of these teams, and a few others, has supported the use of word problem-solving interventions that incorporate explicit strategy instruction, teaching students to identify problem-type information before making a plan to solve, using multiple exemplars, encouraging self-strategies, and using visual representations (Cassel & Reid, 1996; Fuchs, Fuchs, et al., 2008; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al., 2003; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter, 2003; Fuchs, Seethaler, Powell, Fuchs, & Hamlett, 2008; Griffin & Jitendra, 2009; Jitendra, Griffin, et al., 2007; Jitendra et al., 1998; Jitendra & Hoff, 1996; Leh & Jitendra, 2013; Owen & Fuchs, 2002; Seo & Bryant, 2012; Wilson & Sindelar, 1991). For a full review of this literature, including a breakdown of intervention components in each study, please refer to Kingsdorf and Krawec (2016).

Although third grade research has recognized some of the developmental limitations of the larger body of math word problem-solving studies, the researchers still struggled to address appropriate curricular goals. Specifically, the current curricular targets in the studies (e.g. word problems using only addition and subtraction operations) may have matched with the district curriculum at that time, but do not comply with the Common Core State Standards’ expectations of today (e.g. word problems using all four operations). The problem-type strategies employed in all cases were also a mismatch for the Common Core State Standards’ expectations by targeting categories such as buying-bag, pictographs, halves, and difference. The visual representation strategy also raises concerns, as in all cases, the visual representations used were pre-made templates provided by the teachers. Using pre-made templates, depending on the nature of the template, may not allow students to create visual representations that are concrete, aligning with the specific problem information and portraying individualized understanding of the problem. Visual representations certainly permit both concrete and abstract representations; however, third grade students need to move in a developmentally appropriate trajectory from creating individualized and concrete problem representations to the abstract. It is as important for interventions to be developmentally appropriate as it is for them to target the necessary curricular benchmarks that build toward more complex mathematical skills.

Both in mathematics as well as other academic areas, it is empirically supported that early intervention is associated with increased academic success (e.g. Reynolds, Temple, Robertson, & Mann, 2001). While Gersten, Jordan, and Flojo (2005) have recommended the value of early intervention related to mathematical problem solving, very little research has been conducted in this specific area. So, although not yet empirically supported, it seems plausible that successful mathematics word problem-solving interventions, or rather successful mathematics problem-solving teaching procedures, will have greater long-term benefit when applied in third grade, as opposed to middle school. This is likely to increase student success on standardized tests earlier in their academic career, possibly reduce the probability of student identification of LD, and even affect the need for intervention in the later grades.

3. Purpose of the study
It is clear that using explicit instruction with multiple exemplars, while teaching students to identify important problem information for use in visual representation construction, and engaging students in the use of self-strategies, is effective in improving math problem-solving skills. However, applying these intervention components in combination, while using developmentally appropriate curriculum adhering to Common Core State Standards has not yet been investigated. Because of the need for research in this area, the purpose of this study was to evaluate a math word-problem intervention for third graders, which used explicit instruction strategies with multiple exemplars, taught the use of student-generated visual representations, incorporated a self-monitoring checklist, and targeted
Common Core State Standards’ appropriate curricular targets (i.e. all four operations, measurement, estimation of time, masses of objects, and geometric measurement). The specific research questions focused on the individual performance of third graders identified as LD, at-risk, and/or ESOL to determine:

- What was the effect of explicit instruction using multiple exemplars, and a self-monitoring checklist, on increasing correct paraphrasing responses on math word problems?
- What was the effect of explicit instruction using multiple exemplars, and a self-monitoring checklist, on increasing correct visual representation responses in math word problems?
- What was the cumulative effect of this intervention in the target areas of paraphrasing and visualizing on solution accuracy in math word problems?
- Were intervention effects maintained over time?
- How long did intervention implementation and mastery take when presented in a general education inclusive setting?

4. Methods

4.1. Participants and setting

One third-grade Miami-Dade County Public Schools (M-DCPS) classroom participated in the study. The school practiced inclusion but used ability grouping for classroom assignment, a common practice in the M-DCPS district. As a result, the selected third grade classroom was the lowest ability classroom, containing between 15 and 18 students throughout the course of the study (students withdrew and were added to the classroom on a rolling basis). All students belonged to one or more of the following categories (meaning, for example, that a student could be identified as LD and ESOL): LD, at-risk, and English for Speakers of other Languages (ESOL). At the time of identification for these students, the school used the discrepancy model (i.e. a student could be identified as having a learning disability if his or her score on an IQ test was at least two standard deviations higher than his or her achievement test score) to identify students with LD. The school was in the process of transitioning to a Response to Intervention (RTI) model for identification. Students were considered at-risk for LD identification through the RTI model based on a “not-proficient” score in a subject area of math or reading at the start of the year assessment. All of the students identified as at-risk in this classroom scored not proficient in the math assessment area. Not all students in the study were identified as ESOL. However, all students reported speaking another language at home.

Initially, 13 students provided student assent and parent consent forms. Three students withdrew from the school prior to the start of the intervention, which resulted in 10 students being present for the full duration of the study. Specific student information is provided in Table 1. Demographic information regarding race, nationality, and language spoken at home was based on participant self-report. Information regarding age, gender, ethnicity, ESOL level, free/reduced lunch, probability-of-success score on the Florida Assessments for Instruction in Reading (FAIR), retained status, and disability status was taken from school records. The FAIR score (included in the table if students were tested) was included as a measure of student achievement in reading comprehension, since reading comprehension is a part of solving math word problems.

There was one teacher in charge of the classroom. The teacher’s qualifications included: a bachelor’s degree, certification in elementary education and ESOL, native speaker of English and Spanish, and 10 years teaching experience (3 years in the inclusive special education setting). The teacher self-identified as Hispanic. She presented all intervention procedures in the classroom during the mathematics class period; additionally, a special education support teacher and pre-service teacher were occasionally present in the classroom. They were aware of the study procedures and only participated as directed.
4.2. Design

A single-subject design was used to analyze individual student data. Ongoing student progress was tracked with three graphs for each student: one for paraphrasing accuracy, one for visualizing accuracy, and one for computation accuracy. The study was conducted using a multiple baseline across behaviors design (Cooper, Heron, & Heward, 2007).

The two behaviors targeted within the design were paraphrasing accuracy and visualizing accuracy. Baseline data collection on the behaviors of paraphrasing accuracy and visualizing accuracy began for all students simultaneously. Each assessment yielded a score for each behavior at each time point. After steady state responding was achieved for at least 8 of the 10 students (80%) on the paraphrasing behavior, the intervention targeting paraphrasing accuracy began; meanwhile, data continued to be collected on the visualizing behavior, which was still under baseline conditions. After at least eight of the students met criteria on paraphrasing, set at 7/8 across two consecutive sessions, the paraphrasing intervention ended and the visualizing intervention began. The visualizing intervention continued until at least eight of the students met criteria on visualizing, set at 7/8 across two consecutive sessions.

Throughout baseline and intervention, data on computation accuracy were collected. The intervention did not explicitly target computation accuracy, but rather looked at the effects of the intervention with paraphrasing and visualizing on computation accuracy.

4.3. Definition of behaviors

The three dependent variables in the study were paraphrasing, visualizing, and computation accuracy. Paraphrasing accuracy was operationally defined as writing a list of the important information in the problem, including the question (e.g. 36 kids, 1 van = 9 kids, # vans?). Visualizing accuracy was operationally defined as drawing a picture to represent the problem that included all the important information, the question (or the solution to the question), and the operation to be used. Computation accuracy was operationally defined as writing the correct answer to the word problem.

---

Table 1. Student demographic information

<table>
<thead>
<tr>
<th>Student</th>
<th>Age (Years. Months)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Race</th>
<th>Nationality</th>
<th>Language spoken at Home</th>
<th>ESOL level</th>
<th>Identified as Free/ reduced lunch</th>
<th>FAIR score</th>
<th>Retained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>8.4</td>
<td>Male</td>
<td>Not-Hispanic</td>
<td>White</td>
<td>Serbian</td>
<td>Serbian</td>
<td>N/A</td>
<td>At-risk</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>B</td>
<td>8.5</td>
<td>Female</td>
<td>Not-Hispanic</td>
<td>White</td>
<td>Israeli</td>
<td>Hebrew</td>
<td>N/A</td>
<td>At-risk</td>
<td>Yes 64</td>
<td>No</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
<td>Female</td>
<td>Hispanic</td>
<td>White</td>
<td>Argentinian</td>
<td>Spanish</td>
<td>N/A</td>
<td>At-risk</td>
<td>No 55</td>
<td>No</td>
</tr>
<tr>
<td>D</td>
<td>8.11</td>
<td>Male</td>
<td>Not-Hispanic</td>
<td>White</td>
<td>French</td>
<td>French</td>
<td>N/A</td>
<td>At-risk</td>
<td>No 60</td>
<td>No</td>
</tr>
<tr>
<td>L</td>
<td>8.5</td>
<td>Female</td>
<td>Hispanic</td>
<td>White</td>
<td>Argentinian</td>
<td>Spanish</td>
<td>5</td>
<td>At-risk</td>
<td>Yes 79</td>
<td>No</td>
</tr>
<tr>
<td>G</td>
<td>9.7</td>
<td>Male</td>
<td>Hispanic</td>
<td>White</td>
<td>Guatemalan</td>
<td>Spanish</td>
<td>5</td>
<td>Specific Learning Disabled</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>H</td>
<td>9.9</td>
<td>Female</td>
<td>Hispanic</td>
<td>White</td>
<td>Honduran</td>
<td>Spanish</td>
<td>2</td>
<td>Specific Learning Disabled</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>P</td>
<td>8.2</td>
<td>Female</td>
<td>Hispanic</td>
<td>White</td>
<td>American</td>
<td>Spanish</td>
<td>5</td>
<td>At-risk</td>
<td>Yes 43</td>
<td>No</td>
</tr>
<tr>
<td>J</td>
<td>9.7</td>
<td>Female</td>
<td>Hispanic</td>
<td>White</td>
<td>Brazilian</td>
<td>Spanish; Portuguese</td>
<td>2</td>
<td>Intellectually Disabled</td>
<td>No</td>
<td>35</td>
</tr>
<tr>
<td>S</td>
<td>7.9</td>
<td>Female</td>
<td>Hispanic</td>
<td>Asian</td>
<td>Spanish</td>
<td>Spanish</td>
<td>2</td>
<td>None</td>
<td>No 70</td>
<td>No</td>
</tr>
</tbody>
</table>
4.4. Assessments

4.4.1. Problem-solving worksheets
Data were collected throughout the study via permanent products. During each opportunity for assessment, students were presented with a researcher-created double-sided worksheet with two math word problems (e.g. “Our school is taking a class trip tomorrow. Thirty-six students are going. If each van can hold 9 students, how many vans will we need?”). All of the word problems required one step, and included one of the four operations (addition, subtraction, multiplication, and division). The skill focus in each problem aligned with the district pacing guide, which aligned with the math Common Core State Standards (Common Core State Standards Initiative, 2012). Therefore, all problems were within one of the third-grade target areas (i.e. measurement, time, and shape) and included all four operations (addition, subtraction, multiplication, and division). Additionally, to control for operation difficulty and to ensure that each operation was assessed an equal number of times, every two worksheets contained one problem from each operation, with division and multiplication separated across worksheets.

4.4.1.1. Scoring. Under each word problem, there were three headings: paraphrase, visualize, and compute. A paraphrasing response on one word problem was scored out of four. One point was awarded for including each piece of relevant information, including numbers and words/phrases/abbreviations to clarify the relevant information; because each problem contained two pieces of relevant information, this totaled two points. Then, one point was awarded for including the question, and one point for rewording the problem (i.e. maintaining the accuracy of the problem without rewriting it verbatim). This equaled four points for paraphrasing one problem and eight points for paraphrasing each worksheet.

Similarly, a visualizing response on one word problem was scored out of four. One point was awarded for including each piece of relevant information in the picture (totaling two points for the two pieces of relevant information in each problem), one point for accurately placing either the question in the picture (using the written question or a question mark) or the solution to the question if solving had occurred, and one point for selecting the correct operation (i.e. writing the operation symbol). This equaled four points for visualizing one problem and eight points for visualizing each worksheet. For the visualization, the presence of only a number sentence, without labels or the framework of a schematic (e.g. boxes or circles to represent the information components, a number line, or a diagram) did not result in points in the aforementioned categories.

Each word problem was scored out of one for computation accuracy (thus, two points for computation accuracy for each worksheet). Feedback was not provided to the students on their specific problem performance.

4.5. Materials

4.5.1. Student checklist
During lessons, the teacher modeled the use of a researcher-created self-monitoring checklist. The students were allowed access to the checklists during each related intervention phase. An example of this checklist is provided in Appendix A.

4.5.2. Teacher script
The teacher was provided with mock scripts for the initial paraphrasing and visualizing lessons. The scripts provided think-aloud examples, such as the following: “Okay, now that I have read the problem, I am going to paraphrase it, ... I am going to check that off on my checklist.” The teacher was instructed not to read from the scripts during the lessons though, and to refer to them only as examples.
4.5.3. Lesson problems
During each lesson, the teacher used researcher-created word problems that mirrored the structure of the assessment problems. To incorporate the best practice of using multiple exemplars, the teacher used novel word problems when presenting the lessons that included all four arithmetic operations.

4.6. Procedure

4.6.1. Baseline
During baseline, which included a minimum of six baseline measures (12 math word problems) over the course of two months, the teacher presented students with the researcher-created measures in a whole-class setting during regular class time. The students were given 10 min to solve each two-problem measure and were not provided with any assistance or feedback on their performance.

4.6.2. Intervention
The teacher presented the initial explicit instruction lesson on paraphrasing for one class period. The lesson focused on modeling the use of the materials and following a think-aloud procedure to solve a word problem. After modeling, the students practiced with support.

During the second paraphrasing intervention lesson, the teacher started the math period by modeling how to paraphrase a novel word problem, as in the initial lesson. The assessment measure was then presented in the same format as in baseline. The only difference was that students were permitted to use the paraphrasing self-monitoring checklist. As in baseline, the students were instructed to complete all three sections of each problem, including paraphrasing, visualizing, and computation.

This lesson and assessment format, as described on the second day, continued an average of three times a week. As the lessons progressed, the teacher tailored instruction based on assessment data and also modeled more efficient paraphrasing practices (i.e. writing abbreviations of the important information instead of complete sentences) and encouraged the students to do the same. The lessons continued until at least eight of the students met paraphrasing criterion of 7/8 across two consecutive sessions. After paraphrasing was mastered, the visualizing intervention began, which mirrored the paraphrasing intervention procedure.

4.6.3. Follow-up
Approximately 7 weeks after the conclusion of intervention, students were presented with up to eight follow-up assessment measures, administered over the course of 2 weeks. The measure and procedures matched those described in baseline.

4.7. Inter-rater reliability and fidelity
Inter-rater reliability checks were conducted for 20% of each student’s measures at baseline, intervention (paraphrasing and visualizing), and follow-up. The checks were conducted by having a second researcher score the student measure independent of the first researcher. The number of agreements on each measure for each dependent variable were divided by the number of agreements and disagreements, and then multiplied by 100 to yield a percentage agreement score. Any disagreements in scoring were discussed. The average inter-rater reliability agreement score was 95%.

In addition, fidelity checks were conducted on the presentation of the lessons. For the initial lesson for each intervention target and at least once a week thereafter, the researcher observed the teacher presenting the lessons and completed a fidelity checklist (checklist available by contacting the first author). Fidelity checks were conducted for 80% of the paraphrasing lessons (8/10 lessons) and 78% of the visualizing lessons (7/9 lessons). The fidelity percentage score was 100% across all
observations. This score was expected due to the scripted nature of the intervention, and the continued researcher presence and support during the majority of the lessons.

4.8. Social validity
In addition to evaluating the impact of the intervention on the target behaviors, the social acceptability, complexity, and practicality of the intervention, or its social validity (Wolf, 1978), was evaluated. Two researcher-adapted social validity scales were distributed to the students and the teacher at the conclusion of the intervention. Each scale used statements to assess the value of the intervention and its outcomes, as perceived by the students and the teacher (scales available by contacting the first author).

4.9. Graphing and analysis
Individual student data were graphed for paraphrasing, visualizing, and computation accuracy. The paraphrasing and visualizing graphs depicted the probe assessment data as part of the multiple baseline design. Visual analysis was used to examine the data for changes in level, trend, and variability in order to determine if the behaviors changed in a meaningful way, as well as to measure the extent to which the changes could be attributed to the cumulative effect of the intervention. During visual analysis, a decision protocol (Keohane & Greer, 2005) was also used to make moment-to-moment decisions about student progress. Refer to Keohane and Greer (2005) for a full description. The computation graphs depicted the probe assessment data combined within each of the four sections of baseline, paraphrasing intervention, visualizing intervention, and follow-up. Since each assessment was scored out of two for computation accuracy, visual analysis after each probe was not possible (therefore, the visual analysis protocol was only used for the paraphrasing and visualizing data). Additionally, presenting the data cumulatively better enabled an assessment of the overall effect of the intervention within each phase.

In addition to visual analysis, percentages of non-overlapping data (PND; Cohen, 1988) scores were calculated for each student for each main target (paraphrasing and visualizing). PND is a non-parametric approach, which incorporates statistical tests that do not make parametric or distributional assumptions. It is the most widely used effect size measure in single-subject research (Parker & Hagan-Burke, 2007). The PND scores were calculated by counting the number of data points in the intervention phase that did not overlap with the highest data points in the baseline phase, dividing by the total number of data points in the treatment phase, and then multiplying by 100.

5. Results
Overall, the majority of students demonstrated increases in the target behaviors. Individual student graphs and discussion of individual student progress is presented below. PND scores are presented for all students in Table 2. PND scores are interpreted in accordance with the guidelines for interpretation outlined by Scruggs and Mastropieri (1998); PND scores less than 50% reflect an unreliable treatment, PND scores between 50 and 70% reflect a treatment with questionable effectiveness, PND scores between 70 and 90% reflect a fairly effective treatment, and PND scores greater than 90% reflect a highly effective treatment.

5.1. Student Z (At-risk status)
Student Z’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 3. For paraphrasing, his baseline scores were stable at 0 across all seven sessions. Substantial improvements in performance occurred after the introduction of the intervention, with two scores at criterion level (8/8 two consecutive times). His visualizing data were more variable in baseline, but still overall stable, with a mean of 0.85 and a range of 0–2. Again, substantial improvements in performance occurred after the introduction of the intervention, with two scores at criterion level (8/8 two consecutive times). Overall, visual analysis revealed that the intervention proved efficacious for increasing both his paraphrasing and visualizing responses. Furthermore, the overall stable trend in his visualizing data, which demonstrated a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior, provided evidence of a functional relationship
between the intervention and the changes in the target behaviors. As an additional measure of his paraphrasing and visualizing progress, his PND scores are presented in Table 2. His PND scores for both responses were 100%, which represents a highly effective treatment.

Table 2. PND scores for each student and target

<table>
<thead>
<tr>
<th></th>
<th>Paraphrasing (%)</th>
<th>Visualizing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student G</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>Student D</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>Student H</td>
<td>100</td>
<td>38</td>
</tr>
<tr>
<td>Student Z</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Student J</td>
<td>89</td>
<td>50</td>
</tr>
<tr>
<td>Student P</td>
<td>100</td>
<td>67</td>
</tr>
<tr>
<td>Student B</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Student M</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Student L</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Student S</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 3. Student Z’s correct responses to paraphrasing and visualizing rules.
5.2. Student B (At-risk status)
Student B’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 4. For paraphrasing, her baseline scores were stable at 0 across all seven sessions. Substantial improvements in performance occurred after the introduction of the intervention, with two scores at criterion level (8/8 two consecutive times). Her visualizing data were more variable in baseline, but eventually reached a stable 0 trend, with a mean of 0.64 and a range of 0–3. Again, substantial improvements in performance occurred after the introduction of the intervention, reaching criterion level on the second data point with a mean of 7 and a range of 5–8. Overall, visual analysis revealed that the intervention proved efficacious for increasing both her paraphrasing and visualizing responses. Furthermore, the overall stable trend in her visualizing data, which demonstrated a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior, provided evidence of a functional relationship between the intervention and the changes in the target behaviors. As an additional measure of her paraphrasing and visualizing progress, her PND scores are presented in Table 2. Her PND scores for both responses were 100%, representing a highly effective treatment.

5.3. Student D (At-risk status)
Student D’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 5. For paraphrasing, his baseline scores were stable at mostly 0 across the seven sessions,
with one outlying score of 2. This resulted in a mean of 0.29 and a range of 0–2. After the introduction of the intervention, Student D’s scores were still quite variable; this variability coincided with the state testing practice. Over the course of nine sessions, his mean score was 4.56, with a range of 0–8. His visualizing data were also variable in baseline, but stabilized at an overall 0 trend, with a mean of 0.44 and a range of 0–5. In comparison to his established 0 trend in baseline for sessions 6–16, substantial improvements in performance occurred after the introduction of the visualizing intervention, with a mean of 6 and a range of 3–8. Overall, visual analysis revealed variable intervention effects. The changes in the behaviors do seem to be attributable to the intervention though, with the overall stable trend in his baseline visualizing data demonstrating a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior. As an additional measure of his paraphrasing and visualizing progress, his PND scores are presented in Table 2. His PND scores for paraphrasing and visualizing were 56 and 50%, respectively. These scores represent only a questionably effective intervention for Student D.

5.4. Student L (at-risk and ESOL level 5 statuses)
Student L’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 6. For paraphrasing, her baseline scores were variable across seven sessions, with a mean of
1.29 and a range of 0–2. Substantial improvements in performance occurred after the introduction of the intervention, with two scores of 4, followed by two scores at criterion level (8/8 two consecutive times). Her visualizing data were also variable in baseline, but eventually reached a stable 0 trend, with a mean of 0.79 and a range of 0–2. Again, substantial improvements in performance occurred after the introduction of the intervention, with a mean of 5.89 and a range of 3–8. Overall, visual analysis revealed that the intervention proved efficacious for increasing both her paraphrasing and visualizing responses. Furthermore, the overall stable trend in her visualizing data, which demonstrated a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior, provided evidence of a functional relationship between the intervention and the changes in the target behaviors. As an additional measure of her paraphrasing and visualizing progress, her PND scores are presented in Table 2. Her PND scores for both responses were 100%, representing a highly effective treatment.

5.5. Student M (at-risk status)

Student M’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 7. For paraphrasing, her baseline scores were stable at mostly 0 across the seven sessions with one outlying score of 2, resulting in a mean of 0.29 and a range of 0–2. After the introduction of the intervention, Student M’s scores were variable before reaching criterion level. Over the course of seven sessions, her mean score was 6 with a range of 4–8. Her visualizing data were somewhat

![Figure 6. Student L's correct responses to paraphrasing and visualizing rules.](image)
variable in baseline, but hovered around the 0 level most consistently, with a mean of 0.60 and a range of 0–2. Substantial improvements in performance occurred after the introduction of the intervention, with the first two scores reaching criterion level. Overall, visual analysis revealed that the intervention proved efficacious for increasing both her paraphrasing and visualizing responses. Furthermore, the overall stable trend in her visualizing data, which demonstrated a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior, provided evidence of a functional relationship between the intervention and the changes in the target behaviors. As an additional measure of her paraphrasing and visualizing progress, her PND scores are presented in Table 2. Her PND scores for both responses were 100%, representing a highly effective treatment.

5.6. Student G (LD and ESOL level 5 statuses)
Student G's performance on the target behaviors of paraphrasing and visualizing is presented in Figure 8. For paraphrasing, his baseline scores were stable at 0 across all seven sessions. After the introduction of the intervention, Student G did not score at criterion level until his 6th session; his mean score was 3.43 with a range of 0–8. His visualizing data were initially variable in baseline, but stabilized at an overall 0 trend, with a mean of 0.36 and a range of 0–2. He made substantial improvements after the introduction of the visualizing portion of the intervention, scoring at criterion level immediately. Overall, visual analysis revealed variable intervention effects for the paraphrasing
portion of the intervention but strong effects for visualizing. The changes in the behaviors do seem to be attributable to the intervention though, with the overall stable trend in his baseline visualizing data demonstrating a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior. As an additional measure of his paraphrasing and visualizing progress, his PND scores are presented in Table 2. His PND scores for paraphrasing and visualizing were 86 and 100%, respectively. These scores represent a fairly effective intervention for paraphrasing and a highly effective intervention for visualizing.

5.7. Student H (LD and ESOL level 2 statuses)
Student H’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 9. For paraphrasing, her baseline scores were stable at 0 across all seven sessions. After the introduction of the intervention, Student H’s scores reached criterion level immediately. Her visualizing data were somewhat variable in baseline, but hovered around the 0 level most consistently, with a mean of 0.57 and a range of 0–4. After the introduction of the visualizing portion of the intervention, her data, while variable, represented an overall ascending trend. While she did not reach the mastery criterion of two consecutive scores of 7/8 or higher, at the conclusion of the intervention her score was 8/8. The mean and range of her visualizing data in that phase were 3.63 and 0–8, respectively. Overall, visual analysis revealed that the intervention proved efficacious for increasing her paraphrasing responses, with more questionable results for her visualizing responses.
Additionally, the variability in her visualizing data does not provide clear evidence of a strong functional relationship between the intervention and the changes in her behaviors. As an additional measure of her paraphrasing and visualizing progress, her PND scores are presented in Table 2. Her PND scores for paraphrasing and visualizing were 100 and 38%, respectively. This low visualizing PND score was mostly due to an outlying data point in baseline. Considering all data, these scores represent an effective intervention for the paraphrasing response and an unreliable intervention for the visualizing response.

5.8. Student P (at-risk and ESOL level 5 statuses)

Student P’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 10. For paraphrasing, her baseline scores were stable at 0 across all seven sessions. After the introduction of the intervention, Student P’s scores were variable, with an overall ascending trend. She reached criterion level at the 7th session, with an overall mean of 5.38 and a range of 3–8. Her visualizing data were variable in baseline, with a mean of 0.93 and a range of 0–4. After the introduction of the visualizing portion of the intervention, she scored at criterion level quickly (on the second session). Overall, visual analysis revealed that the intervention proved efficacious for increasing her paraphrasing and visualizing responses. However, a functional relationship was only weakly demonstrated due to the variability in her visualizing data. As an additional measure of her paraphrasing
and visualizing progress, her PND scores are presented in Table 2. Her PND scores for paraphrasing and visualizing were 100 and 67%, respectively. These scores represent an effective intervention for the paraphrasing response and a questionable intervention for the visualizing response.

5.9. Student S (ESOL level 2 status)
Student S’s performance on the target behaviors of paraphrasing and visualizing is presented in Figure 11. For paraphrasing, her baseline scores were stable at 0 across all five sessions. After the introduction of the intervention, Student S’s scores were variable before meeting criterion. Over the course of seven sessions, her mean score was 6.57, with a range of 2–8. Her visualizing data were mostly stable in baseline, hovering around the 0 level most consistently, with a mean of 0.46 and a range of 0–2. Substantial improvements in performance occurred after the introduction of the intervention, with the second score reaching criterion level. Overall, visual analysis revealed that the intervention proved efficacious for increasing both her paraphrasing and visualizing responses. Furthermore, the overall stable trend in her visualizing data, which demonstrated a lack of improvement in visualizing despite the introduction of the intervention for the paraphrasing behavior, provided evidence of a functional relationship between the intervention and the changes in the target behaviors. As an additional measure of her paraphrasing and visualizing progress, her PND scores are presented in Table 2. Her PND scores for both responses were 100%, representing a highly effective treatment.
5.10. Cumulative computation performance
The cumulative computation performance of all students is presented in Table 3. For Students D, Z, M, and S computation accuracy improved throughout the duration of the intervention, and further increased at follow-up. For Students G, H, J, and P, computation accuracy at follow-up was an overall improvement, in comparison to baseline. Students B and L already had high computation accuracy skills in baseline.

5.11. Social validity
The students’ responses to the intervention were mostly positive. Question 1 addressed problem solving in general, with the majority of students feeling that problem solving is important (70%). Questions 2 and 6 focused on the skill of paraphrasing, and questions 3 and 7 focused on the skill of visualizing. Overall, the students appeared to value the intervention’s visualizing component more than the paraphrasing component, with a 100% favorable response to both visualizing questions and more variable responses for the paraphrasing questions. Questions 4, 5, and 8 inquired about the overall intervention. The responses were somewhat variable, but overall students felt that the intervention strategies were now automatic (70%), will continue to be used (80%), and enhanced problem-solving ability (80%). However, these social validity results should be interpreted with caution. Having the teacher interview the students about what is essentially her teaching effectiveness is likely to impact the validity of their responses, as young students in particular are susceptible to
social desirability bias; that is, their relationship with the teacher may hinder their ability to truthfully express an opinion about the intervention’s value.

The teacher’s responses were also positive. She strongly agreed or agreed with all 16 questions assessing the value of the intervention, 69 and 31%, respectively.

6. Discussion
Overall, the multi-component intervention proved effective in increasing the paraphrasing, visualizing, and computation responses for all students. The main intervention components were the use of explicit instruction, a self-monitoring checklist, and multiple exemplars in instruction and assessment practices. However, due to the treatment package nature of the study, determining the most valuable pieces of the intervention is difficult.

The explicit instruction component of the intervention was present in all 19 instructional sessions. It provided the overall structure of the intervention format (i.e. guiding the teacher’s lesson structure, use of modeling, and opportunities for student practice and feedback). In the work of Fuchs and colleagues (e.g. Fuchs, Fuchs, et al., 2008; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al., 2003; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter, 2003; Fuchs, Seethaler, et al., 2008), an explicit instruction lesson format was found to be superior to a general instruction format when delivering an intervention which increased the correct problem-solving responses of third graders. Jitendra and colleagues (e.g. Jitendra, Griffin, et al., 2007; Jitendra et al., 1998; Jitendra & Hoff, 1996) also found that using an explicit instruction format was most beneficial when implementing a visualization intervention with a similar population. Both researchers utilized interventions that included multiple components but identified explicit instruction as a necessary piece. Therefore, that this intervention was delivered in an explicit instruction format and resulted in favorable student outcomes aligns with the previous research on the critical characteristics of effective math problem-solving interventions for third graders.

Separating the explicit instruction component and the instructional guidance provided with the self-monitoring checklist is impossible, as the teacher consistently modeled use of the self-monitoring checklist during the lessons, intertwining it with the explicit instruction strategy. The teacher used the checklist to drive her explicit instruction, and that was where the checklist played the most critical role—during instruction. During student assessments, the students were rarely observed using the self-monitoring checklists during solving. However, although the students were not observed reading and checking off the items on the checklist on a regular basis, inspection of student work after assessment revealed that the behaviors outlined on the checklist were being completed (e.g.

<table>
<thead>
<tr>
<th>Table 3. Cumulative percentage correct computation points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>J</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>S</td>
</tr>
</tbody>
</table>


underlining the important information). The work of Marsico (1998) on increasing the independent math performance of students diagnosed with various learning difficulties found that increases in their correct responses were the observable result of students’ self-monitoring (self-editing) behavior when using a checklist. Harris and colleagues (e.g. Harris, Danoff Friedlander, Saddler, Frizzelle, & Graham, 2005) discussed similar results when investigating the effects of self-monitoring on the academic performance of students with ADHD and finding that the intervention controlled independent on-task spelling behaviors. Like results were found here, with the behaviors on the checklist becoming automatic and happening covertly, but the checklist still playing a role in increasing independent student responding. As in the work by Marsico (1998) and Harris and colleagues (e.g. Harris et al., 2005), the role of the checklist became one of a discriminative stimulus (SD) for independent problem solving, rather than a tool needed to facilitate each step in a problem-solving algorithm. That is, the response of solving in the presence of the checklist during the instructional sessions resulted in praise (an established reinforcer for the students), so during assessments the students responded to its presence by again solving. The checklist therefore became the cue to solve the problem, regardless of the need to read and follow the individual steps on the checklist. This role of the checklist aligned with the students’ social validity data.

Only 70% of the students reported no longer needing the self-monitoring checklist. The checklist did serve a function for the students, even if they were not observed reading and checking it off.

Using multiple exemplars is good practice and has proved effective in increasing the generalization responses in problem-solving intervention studies by Fuchs and colleagues (Fuchs, Fuchs, et al., 2008; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al., 2003; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroe, 2003; Fuchs, Seethaler, et al., 2008; Owen & Fuchs, 2002). Problem-solving work by Hughes (1992) also supported this, finding that the use of multiple exemplars in an instructional package resulted in not only response generalization, but also response maintenance behaviors. All assessments in this study were presented in a similar context; explicit tests for generalization did not occur. However, the teacher did report that the students were using the paraphrasing and visualizing strategies during their other assignments. This generalization may be due to the use of multiple exemplars in the teaching practices. Additionally, while tests were not conducted for response generalization across settings, in the follow-up procedures, probes were conducted for response generalization across time. As in Hughes (1992) data, the follow-up data here support the use of the multiple exemplars in maintaining the computation skills of the students.

### 6.1. Limitations

While the intervention proved valuable, three main limitations exist. First, because most students belonged to multiple groups, trying to draw conclusions of intervention effectiveness based on student classification status was difficult. This may limit the generalizability of the intervention for specific student groups. However, it provides support for the single-subject design and analysis used in this study that focused on individual student data.

Second, although the design and analysis used within this study were the most beneficial for exploring the results in relation to unique student characteristics, certain environmental constraints limited the use of best design practices. While each student did serve as his or her own control within the study, further measures would have been beneficial to increase the validity of the design. For example, putting the students into two groups where the intervention phases were counterbalanced would have provided more information about the role of paraphrasing and visualizing skills within the problem-solving process, and also would have addressed the potential confounds of sequence effects. This was not possible though, since all students were part of the same class, and the intervention was delivered in a group lesson format. Furthermore, the theory behind mathematical problem solving places the process of paraphrasing before visualization. Therefore, while counterbalancing processes would have strengthened the methodology of the study, it may have done so at the expense of the intervention. Additionally, two of the students, Student B and Student L, had high computation scores at baseline that decreased after intervention. The decision to include these
students in the study could be seen as a limitation. However, a few things are worth noting. First, although decreases in computation scores occurred for both students, their scores were still relatively high. Student B's scores increased in the visualization phase to 93% before decreasing in the follow-up phase to 75%. Student L’s score was 69% at follow-up. These decreases were attributable to discrete issues for both students. Student B had limited opportunities for assessment because of excessive absences. Student L routinely ran out of time during assessment before she was able to complete computation. Second, although both students were proficient in computation, their visualization scores were very low at baseline. The intervention improved visualization abilities in both students that may not be needed with the more simple third grade problems presented here, but will likely help them as problems increase in complexity. As related to this, both students were observed using these visualization skills outside of the intervention context with more complex problems. Although the link between their visualization and computation abilities was not represented in the data here, more practice in fluency for Student L and more opportunities for assessment with more complex problems for Student B may reveal otherwise.

Third, although the intervention was teacher-delivered, the researcher was present during the majority of the lessons. It is likely that reactivity occurred; that is, the researcher’s presence may have had an impact on the behavior of the teacher and the students. It would be valuable to further assess the ecological validity of the intervention by having it delivered in a more natural context by the teacher without the presence or support of the researcher.

A few other limitations that were out of the control of the researcher are also worth noting. One area of difficulty was that interim testing (i.e. testing used to prepare for the upcoming standardized state assessment) occurred during the first week of the paraphrasing intervention. This resulted in an interruption in the intervention schedule and a significant disruption to the daily schedule. This disruption may have been reflected in the paraphrasing score drop seen in the data for some students (e.g. Student J, Student S, Student G, Student M, and Student D) after the first paraphrasing assessment within the intervention phase. Second, the transient nature of the school’s population resulted in numerous changes to the classroom during the study. Students withdrawing from the school resulted in attrition of three participants. There were also several students who either became part of the class or were moved to a different classroom during the course of the study. This disrupted the general flow of classroom operations, like the structure of the teacher’s lessons (e.g. having one student translate her instruction into Russian for a new Russian-only speaking student) and the classroom seating arrangement.

6.2. Educational significance

Despite these limitations, the results of this study offer valuable contributions to the field. Previous research with this age group (e.g. Fuchs, Fuchs, et al., 2008; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al., 2003; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter, 2003; Fuchs, Seethaler, et. al., 2008) has focused on the mainly cognitive aspects of the mathematical problem-solving process. Within this study a cognitive and behavioral model of mathematical problem solving was developed to better understand the observable skills that may demonstrate student understanding in this complicated process. Looking at the discrete skills associated with paraphrasing and visualizing and investigating their impact on computation were areas not previously explored; in doing this, this study generated some conclusions as well as important questions for future investigation. It also provided teachers with a different way of teaching and assessing problem solving in the general education curriculum. Teachers are likely to feel more comfortable teaching and assessing skills of the problem-solving process which are observable (e.g. underlining important information, schematically representing parts of the problem) to make determinations about student areas of need, their own teaching effectiveness, and student problem-solving ability, as opposed to trying to draw conclusions based on unobservable cognitive processes or student answers alone. Such a process of student learning and teaching evaluation aligns with the current RTI framework being implemented in many schools across the nation. A basic RTI framework includes ongoing tracking of student progress through stages of prevention and intervention, which
includes: (a) student screening, (b) implementing evidence-based intervention, (c) monitoring student progress, and (d) analyzing student progress to determine special education eligibility (Fuchs, Mock, Morgan, & Young, 2003). The intervention applied here began with a baseline assessment of problem-solving skills (stage a), implemented evidence-based practices of using explicit instruction, multiple exemplars, and self-strategies (stage b), monitored student progress through ongoing assessment of problem-solving skills (stage c), and examined student learning through the graphing and analysis of individual student problem-solving data (stage d). Students identified in stage d as struggling with problem solving despite the application of high-quality instruction at stage b, may then benefit from more intensive, individualized, one-on-one problem-solving practice using the strategies applied here (e.g. working with a special educator who systematically teaches the student each step of creating a schematic representation to mastery criterion). This model could seamlessly be applied by educators committed to using evidence-based mathematical teaching practices in the classroom to make determinations about student classification and remediate learning difficulties. This practice appeared to resonate with the teacher in this study, based on her very favorable feedback on the social validity scale and request to apply these teaching procedures and assessments with her future classes.

Furthermore, previous research on math problem-solving interventions at the third grade level failed to address critical issues such as incorporating curriculum which aligned with the Common Core State Standards, targeting all four operations, targeting practice and assessment with a broad set of word problems, and teaching visualization strategies which were not problem-type specific. These concerns, addressed through this intervention, make the intervention practices applied here more aligned with the regular education curriculum and resultanty more valuable in the general education inclusive classroom and for the general education teacher. Also, the overall format of the intervention employed a very natural and simple framework. The teacher required very little training to learn the teaching strategies used here and felt flexible to make the intervention her own (i.e. was not required to use a script or a set of contrived practice and assessment materials). These aspects increase the likelihood of intervention use outside of the intervention context; in fact, the teacher in this study reported that she used the intervention strategies in her teaching practices outside of the intervention sessions, and will continue to incorporate the strategies in upcoming years.

Lastly, the majority of previous research that investigated related intervention strategies presented in a group inclusive setting did not sufficiently disaggregate and discuss individual student data. As a result, the effects of the interventions on students with special education needs were often masked. In cases where student data were broadly categorized by disability status, as in the study by Owen and Fuchs (2002), the accuracy of post-test scores for students with disabilities only reached 45%, a still failing level. In this study, 7 of the 10 students reached a computation accuracy level at follow-up of about 70% or higher. Overall, this study was able to show that students with LD, ESOL classification, and at-risk status were all able to benefit from the intervention in a practical way by increasing their problem-solving skill set. This is useful when considering the current service delivery model of special education services and the variety of student needs that a teacher in this context strives to meet.

Acknowledgements
The authors are grateful to the students as well as the teachers, administrators, and other school personnel in the Miami-Dade County Public Schools for their cooperation and support.

Funding
The authors received no direct funding for this research.

Author details
Sheri Kingsdorf1
E-mail: sheri.kingsdorf@bcbahours.com

Jennifer Krawec2
E-mail: krawec@miami.edu
1 BCBAhours, 401 Cheyenne Blvd, Colorado Springs, CO 80905, USA.
2 School of Education and Human Development, University of Miami, 1507 Levante Ave, Coral Gables, FL 33146, USA.

Citation information
References


Appendix A

Sample student checklist

Paraphrase

SAY: Underline the important information.
ASK: Have I underlined the important information?
SAY: Put the problem in your own words and write the important information.
CHECK: That the information goes with the problem.

Paraphrasing prompt
I’m trying to figure out ____________. I know that ____________ and that ____________. Again, I’m looking for ____________.