RTI IN A MIDDLE SCHOOL: FINDINGS AND PRACTICAL IMPLICATIONS OF A TIER 2 READING COMPREHENSION STUDY

Michael Faggella-Luby and Michelle Wardwell

Abstract. Response to intervention (RTI) has received considerable attention from both researchers and practitioners as a schoolwide model for service delivery. However, research is limited on RTI applications in middle and high schools. The purpose of this article is to describe the outcomes of an experimental examination of a secondary (Tier 2) literacy intervention for at-risk fifth- and sixth-grade students in an urban middle school assigned to one of three conditions: Story Structure (SS), Typical Practice delivered by reading specialists (TP), and Sustained Silent Reading (SSR). Results indicated a statistically significant difference between the mean posttest cloze scores of the SSR group and both the SS and TP conditions. Study findings support the growing body of research indicating that at-risk students need intensive and explicit instruction in addition to opportunities to practice reading.

MICHAEL FAGGELLA-LUBY, Ph.D., Center for Behavioral Education and Research, University of Connecticut, Storrs, Connecticut.
MICHICHEL WARDWELL, Ph.D., Center for Behavioral Education and Research, University of Connecticut, Storrs, Connecticut.

Concerns about improving literacy-related outcomes for middle and high school students are providing a catalyst for reconceptualizing service delivery for at-risk students. In recent years, a multitiered model, response to intervention (RTI), has received considerable attention from both researchers and practitioners as a schoolwide model for improving service delivery for all students (Fuchs, Mock, Morgan, & Young, 2003; Graner, Faggella-Luby, & Fritschmann, 2005). Consequently, RTI models are under rapid development and adoption across the United States. However, the empirical foundations of RTI are rooted in early-literacy research and the elementary school context, causing practical challenges when the model is applied to middle school settings. Specifically, schools face the initial challenge of differentiating Tier 2 instruction from Tier 1 by considering when and how to deliver supplemental instruction, and which practitioners are qualified and available to deliver intensive instruction. Therefore, empirical study of RTI in middle schools is warranted to validate the efficacy of implementation efforts.

This article describes the findings and practical implications of an experimental examination of three Tier 2 reading comprehension interventions for at-risk fifth- and sixth-grade students in an urban middle school.

A Significant Need
Recent National Assessment of Educational Progress (NAEP) data indicate that 68% of fourth graders and 70% of eighth graders in public schools perform at or below the basic level in reading comprehension nation-
ally (National Center for Education Statistics, 2009). Operationally, these students may struggle with recognizing literal information, explicitly stated reasons for character action, or finding the central problem or main idea from text. In the state of Connecticut, where the present study was conducted, NAEP data demonstrate a trend similar to the national average. However, there are no states with a larger reading achievement gap (comparing percentages of students within subgroups that are at or above proficiency and those who are not) between students receiving free and reduced-price lunch and those who are not, white and black students, and white and Hispanic students at the fourth and the eighth grade (State Department of Education, Connecticut, 2007). Accordingly, schools in Connecticut and around the nation are in search of a new way to conceptualize service delivery that results in improved outcomes for all students.

Response to Intervention as Reform Model

One promising schoolwide model for improving student outcomes, RTI, is a multitiier model of schoolwide service delivery intended to address the needs of academically diverse groups of students by measuring their response to research-based instruction (Fuchs et al., 2003; Graner et al., 2005). RTI's empirical foundations are rooted in several traditions, including early-literacy research (K-3), and is viewed as an alternative procedure in identification of specific learning disabilities (Simonsen, et al., 2010).

The model is generally operationalized in one of two configurations: (a) as problem solving, in which a team of practitioners uses a recursive process to determine necessary instruction based on reasons for under-achievement via a case-by-case analysis; or (b) as standard treatment protocol (STP), in which struggling students receive an initial standard intervention as a means to prevent failure (Graner et al., 2005).

RTI is typically delivered via three tiers of increasingly intense instruction with varied duration, time, and frequency of intervention (Mellard, 2004; Troia, 2005). While each state may adopt its own version of RTI (e.g., the SRBI model in Connecticut; see Spear-Swerling, 2008), the model is distinguished by eight core features and five implementation attributes (Mellard; Troia). Most notable among these characteristics are: (a) instructional practices that are evidence-based; (b) progress monitoring to establish rate and level of student learning (i.e., measuring student response to instruction); (c) an assessment of fidelity of implementation; and (d) universal screening.

Regardless of the specific RTI model implemented, considerable questions remain about the validity of applying RTI beyond elementary school. Summarized in Fuchs and Deshler (2007), such questions include (a) the relevance of RTI to middle and high schools; (b) components of effective multitiered instruction that address curricular demands (e.g., reading comprehension, writing, content area knowledge); (c) who will carry out the instruction across the multiple tiers (and necessary areas of expertise); and (d) how to routinely measure student progress and teacher fidelity. In addition to these important issues, of primary importance for practitioners is determining what the focus, or content, of supplemental instruction shall be in middle schools.

Content of Supplemental Instruction

Fortunately, during the last decade, a converging, yet multifaceted, evidence base has emerged regarding reading comprehension instruction in grades 4-12. Faggella-Luby and Deshler (2008) analyzed six comprehensive literature reviews to provide a basic foundation for making informed programming and policy decisions for middle and high school students, especially those at risk for failure. A key finding from these reviews is that reading comprehension for students at risk for failure as well as their typically achieving peers can be improved with targeted instruction on what good readers do (Faggella-Luby & Deshler). Research related to three potential options for supplemental instruction in middle schools is detailed below.

A Content Area-Related Strategy. Struggling readers require robust evidence-based interventions during the transitional years of middle school (grades 5 through 8) to remediate learning difficulties and to prepare students to meet the demands of the middle school curriculum (Faggella-Luby & Deshler, 2008). Moreover, given that previous learning for struggling readers may have been incomplete, preventing mastery of critical content-area knowledge and advanced skills, interventions that embed instruction with content-area learning may prove doubly effective. For example, supplemental instruction in the features of story structure is not only part of typical English/language arts standards-based curriculum (e.g., CT State Standards, www.sde.ct.gov), but is also supported (a) theoretically to improve reading comprehension by providing students with a macrostructure to organize essential content from narrative passages (e.g., Kintsch, 2004); and (b) empirically to improve basic academic performance and lead to higher-order thinking, including causal reasoning (Gersten, Fuchs, Williams, & Baker, 2001).

Three successive phases of research over the last 30 years lend empirical support to story structure instruction (Faggella-Luby, Schumaker, & Deshler, 2007; Olson & Gee, 1988). In particular, studies have demonstrated improved reading comprehension outcomes for at-risk
middle and high school students when applying a variety of story structure-related strategies (e.g., Dimino, Gersten, Carnine, & Blake, 1990; Faggella-Luby et al.; Gardill & Jitendra, 1999). Additionally, in the studies by both Dimino et al. and Faggella-Luby et al., students who scored low on pretests did not differ statistically from their high-achieving peers on the posttest. This is significant, because one of the stated goals of RTI is to close the gap between students with disabilities or at risk for failure and their typically achieving peers.

**Generic Components of Comprehension Instruction.** While the above strategy instruction is related to content-area knowledge specifically, there is also empirical evidence to support the use of more generic components of comprehension instruction. Consistent with the recommendations set forth by the National Reading Panel (NRP) of the National Institute of Child Health and Human Development (NICHD; 2000) and meta-analyses on strategy instruction (e.g., Swanson, 1999), two recent reviews related to readers in grades 4-12 underscore the importance of certain types of comprehension instruction. For example, Faggella-Luby and colleagues (e.g., Faggella-Luby & Deshler, 2008; Faggella-Luby, Ware, & Capozzoli, 2010) identified teaching essential content and vocabulary, individual cognitive strategies (e.g., self-questioning, summarizing or self-monitoring) or packages of strategies (e.g., Reciprocal Teaching; Palincsar & Brown, 1984), improving student motivation and engagement, and activating prior knowledge as critical to comprehension for struggling readers. Additionally, the NRP demonstrated mixed support for the benefits of guided reading (NICHD). Consequently, any of the above practices might be considered justified as part of typical practice in middle school supportive learning environments in an effort to improve reading comprehension outcomes for struggling readers.

**Sustained Silent Reading.** A potential challenge to improving literacy instruction using the two methods listed above is the requirement for practitioner expertise in either content-area knowledge (i.e., highly qualified history teachers are more likely to embed strategies that access critical content than support educators who do not have the content area background) or reading instruction (i.e., reading specialists have rich background knowledge in reading instruction that content area teachers in middle schools often lack). Therefore, given the variability of expertise, and the likely costs of professional development, a potential standard treatment for all struggling readers is to build time for reading practice that is practical, cost efficient, and considered to be a beneficial practice correlated with reading achievement (NICHD, 2000). Moreover, programs like SSR enjoy widespread practitioner support, allowing for higher likelihood of implementation.

**Method**

**Purpose and Design of the Study**

Given the pressing need for empirical study of RTI models beyond elementary school and implementation of research related to the content of reading comprehension instruction for at-risk middle school readers, the following research study was designed. Specifically, the purpose of this study was to investigate the effects of three standard treatment conditions of instruction, Story Structure (SS), Typical Practice (TP), and Sustained Silent Reading (SSR), by randomly assigning at-risk fifth- and sixth-grade students to each condition.

Due to study constraints, a posttest-only design was used. Though considered a robust design (Shadish, Cook, & Campbell, 2002), the following steps were taken to strengthen the design: (a) random assignment of the subject population (with matching explained below), (b) use of proxy pretest data to confirm similarities between groups, (c) comparison of instruction to typical practice, and (d) multiple posttests (including standardized measures; see Shadish et al.).

**Participants**

Participants were five female middle school teachers and 86 fifth- and sixth-grade at-risk students attending a public urban middle school in a northeastern state. The two SS group teachers were fifth-year student interns working on their master’s degrees in education at a local university. The three typical practice TP group teachers were certified reading specialists with master’s degrees and 9, 29, and 29.5 years of experience, respectively.

Student participants were selected for their at-risk status, determined by analysis of reading comprehension ability as measured on the Degrees of Reading Progress (DRP) screening test scores. The DRP is a nationally norm-referenced measure of reading used to determine how well students understand and process text, and is used as part of the Connecticut Mastery Test, which is the high-stakes test for Connecticut (Connecticut State Board of Education [CSBE], 2007).

Students were selected for the study if their scores fell below the criterion of 48 for fifth graders and 52 for sixth graders, identifying them as at risk, or students who can comprehend, with varying degrees of difficulty, materials below a fifth- or sixth-grade level, respectively (CSBE, 2007). Selected fifth and sixth
graders fell within the score range of 32 to 44, and 36 to 52, respectively (see Table 1 for student demographic data). Students with documented disabilities in reading were excluded from selection at the school’s request as they were already receiving supplemental instruction.

Once selected, students were randomly assigned to one of three conditions: (a) experimental (SS), (b) comparison (TP), and (c) SSR (conditions are described below). One-way ANOVAs confirmed no statistical differences between groups on either DRP placement data, \( F(2, 76) = .134, p = .875 \), or school-collected fall lexile data, \( F(2, 84) = .386, p = .681 \). Attrition for the SS group was 13% (4 students), for the TP group, 7% (3 students), and for the SSR group, 18% (4 students). Although students were randomly assigned to condition, group sample sizes differed due to attrition, failure to obtain signed permission forms, and the fact that the number of teachers per condition differed (i.e., 2 SS, 3 TP, 2 SSR).

**Setting**

The study took place in an urban middle school serving a population of 986 students in grades 5-8. During the previous school year, 68.2% of the students were eligible to receive free/reduced-price meals, and 30.8% were listed as having a non-English home language. The school population consisted of 66.9% students of color, with 60.2% Hispanic, 33.1% White, 4.9% Black, 1.3% Asian American, and 0.5% American Indian.

Supplemental, Tier 2, instruction was provided in addition to the traditional English/language arts (ELA) Harcourt Brace curriculum in separate (non-ELA) classrooms during the first and second quarters of the school year. Classroom setup varied according to teacher; however, all instruction took place within small classrooms fitted with a chalk or whiteboard for writing and desks or tables to accommodate up to 12 students.

**Instructional Materials**

Passages of varying levels of difficulty, including Aesop’s Fables, folktales, and high-interest/low-vocabulary (e.g., Hiebert & Fisher, 2005) level texts were used during instruction and student practice for the SS condition. Late in the study, the teachers in the SS condition gained access to the reading specialists’ classroom libraries, and subsequent instruction included those resources (more high-low texts). The TP teachers used books from their classroom libraries for instruction and student practice. SSR students chose books from the school library, brought books from home, or were provided books by the teacher if they forgot their own.

**Instructional Conditions**

**SS Condition.** Experimental SS instruction employed a modified version of the Embedded Story-Structure ESS Routine (Faggella-Luby et al., 2007). Specifically, the SS intervention teaches three strategies to improve reading comprehension. First, students learn to ask themselves seven story-structure-related questions (self-questioning) based on canonical story-structure elements (and ELA curriculum literary terms), including main character, initiating event, time, place, central conflict, climax/turning point, resolution, and theme (e.g., Stein & Glenn, 1978). Second, students engage in story-structure analysis by identifying and labeling specific elements from the narrative on a Story-Structure Diagram. Students label the diagram after they have read the text and answered the self-questions. Third, students learn to use a five-sentence summary writing formula to produce a written account of the narrative that includes the critical canonical components. Students use a graphic organizer to record self-questions (and answers), the SS diagram, and written summary.

SS instruction was modified from the original ESS Routine in both its content and instructional methodology to fit setting constraints related to instructional time and teacher preference. Specifically, the content was modified by (a) removing explicit teaching of strategy-related mnemonics (instead, mnemonics were placed on posters in the classroom); and (b) an additional sentence was included in the summary formula. The instructional methodology was modified substantially as in the current study, (a) instruction took place in 30-minute sessions 2-3 days per week over 18 weeks instead of over 9 days in 2-hour blocks, (b) no cooperative learning was used to mediate student learning, (c) student opportunities to practice and receive corrective feedback were considerably reduced, and (d) significantly more time was spent during instruction on teacher description and modeling of SS self-questions and diagram than on student independent practice.

In this condition, the three SS components were initially introduced, described, and modeled by the teacher during the first week of instruction. This expert teacher model provided a “big picture” example of how the targeted strategies are used effectively and served as a referent when teaching individual strategy components. Of the three components, self-questioning instruction required significantly more time than the other two components, as the literary terms (e.g., central conflict, initiating event) were simultaneously defined and examples from short passages (e.g., Aesop’s Fables) were provided. Of the 18 weeks of instruction, approximately 9 weeks focused on defining terms and student practice of self-questioning, 2 weeks focused on story structure analysis, and 2 weeks focused on summarizing. The remaining five weeks of instruction were spent on student independent practice of the entire SS Routine. Teachers followed a written protocol for each lesson, and concluded each class with a classwide review.
of the day's lesson. Students handed in completed graphic organizers periodically for corrective feedback.

**TP Condition.** Typical practice instruction mirrored the individual beliefs and philosophies of the three reading specialists who delivered the instruction when attempting to improve reading comprehension, and generally fell within guidelines outlined by the NRP (NICHD, 2000). Each reading specialist independently designed instruction, without input from the research team, to allow for comparison of the SS to alternative typical-practice Tier 2 instruction. Documentation of TP was achieved through regular fidelity observations (see below).

Though varied, instruction was characterized by four regular activities across the three TP groups. Primary instruction focused on mini-lessons on the components of active reading, including preview, prediction, character identification, summarizing, visualizing, and questioning. Each mini-lesson consisted of a brief introduction of the term by the teacher (3-7 minutes) followed by a guided reading activity in which students were prompted by the teacher to use the new skill. Students were given an Active Reader Card that provided a visual mnemonic to help remember the components. The card included a list of the components covered in the mini-lessons to prompt student use. Second, students were taught vocabulary that corresponded to the story they were reading and consisted of both definitions and examples. Third, students were regularly engaged in guided reading through the use of Literature Circles (Daniels, 1994) or independent reading to practice the Active Reading skills. Finally, a significant amount of time was spent on journal writing. Journaling took the form of reader response to text or answering teacher-initiated questions.

Instructionally, TP lessons began by referencing lesson objectives that were written on the board. Mediation of skill learning followed an implicit methodology, typically using the mini-lessons to describe a skill and provide opportunities for discussion, student questions, and independent practice. Terms were rarely defined or modeled by the reading specialists using think-aloud problem solving. Definitions were considered sufficient to result in generalized use.

**SSR Condition.** SSR students were engaged in Sustained Silent Reading (SSR) each day of intervention instruction for 30 minutes each. At the start of SSR time, all other work was put away. Students sat alone or at small tables of two to four students reading silently. Occasionally, students put their heads down. When this happened, they were instructed by teachers to keep reading. Students were encouraged to bring a book to class for use during SSR time and were not allowed to work on other assignments. When students did not bring books to class, one was provided for them from a classroom library. Classrooms were kept quiet with minimal distractions to allow for silent reading. Teachers regularly engaged in reading during this time as well.

**Instructional Procedures**

Teacher-to-student instructional ratios were 1:5-7, depending on class enrollment and attrition for each grade level. Instruction for each class took place 2-3 days per week in 30-minute sessions during the first 2 quarters of the school year. Typically, variations in weekly meetings were a result of school scheduling and were the same for students in all groups. However, during the university semester break, SS teachers missed four weeks of teaching. To mediate learning during this period, students were prepared to complete independent practice with texts of their own choosing. Students were asked to complete two SS organizers, which the teachers collected upon returning to teaching. In total, SS received 35 days of instruction, whereas TP and SSR students received 48 days.

**Dependent Measures**

**Cloze.** The AIMSweb Maze (Shinn & Shinn, 2003), a standardized, curriculum-based cloze measure, was administered to all students at posttest. The Cloze is a timed, multiple-choice assessment that measures sentence-level reading comprehension. Students silently read an approximately 375-word passage in which every seventh word has been replaced with three words in parentheses; one of the three choices is the correct word from the passage. Students have three minutes to complete as many items as possible. To control for variations in passage-level difficulty (e.g., Oakland & Lane, 2004), a counterbalanced design using fifth-grade passages was administered to all students. For counterbalancing at posttest, half of the class was administered Form A; the other half received Form B. The Cloze score was determined by adding up the total number of correct answers.

**Strategy-Use Test.** The Strategy-Use Test was administered to all students at posttest. This measure was used to assess the degree to which the experimental students learned the SS strategies taught in the intervention (which the comparison students did not receive). The evidence provided from this measure could later be compared to student performance on reading comprehension measures to provide additional information related to strategy acquisition and improvements in comprehension.

The test comprised three sections, each pertaining to one of the SS strategies. In the first section, seven lines were provided for students to write prereading questions related to novel passages. The second section
Table 1

**Student Demographics**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Story Structure</th>
<th>Typical Practice</th>
<th>Sustained Silent Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5th</td>
<td>6th</td>
<td>5th</td>
</tr>
<tr>
<td>Mean Age (in months)</td>
<td>132</td>
<td>143</td>
<td>131</td>
</tr>
<tr>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>58</td>
<td>4</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Caucasian</td>
<td>2</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9</td>
<td>75</td>
<td>9</td>
</tr>
<tr>
<td>Other (Ukrainian)</td>
<td>1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Disability Status</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ELL Status</td>
<td>2</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Free/Reduced-Price Lunch</td>
<td>8</td>
<td>67</td>
<td>13</td>
</tr>
<tr>
<td>DRP Unit Score M(SD)</td>
<td>36.9 (3.06)</td>
<td>43 (4.40)</td>
<td>35.6 (2.93)</td>
</tr>
</tbody>
</table>

asked students to correctly label a Story-Structure Diagram based on a 375-word narrative fifth-grade passage from AIMSweb. For the third section, students were prompted to write a 60-word summary of the passage. Passages were chosen based on their inclusion of each of the eight elements of narrative text structure and were not used in the Cloze activity.

Similar to the Cloze assessment, a counterbalanced design using fifth-grade passages for all students was used to control for passage-level difficulty across measures. Again, for counterbalancing, Form A was administered to half the class whereas the other half of the students completed Form B. The test was scored section by section using an answer key. The total number of points possible was 25, with eight points available in both the prereading questions and the summary, and nine points available for correctly drawing and labeling the diagram.

**Gates-MacGinitie Reading Comprehension.** To further assess reading achievement, the Comprehension portion of the Gates-MacGinitie Reading Test, Fourth Edition (GMRT-4), was administered to all groups (MacGinitie, MacGinitie, Maria, & Dreyer, 2002). The GMRT-4 is a standardized, 35-minute reading measure in which students answer multiple-choice questions related to text passages. Students were tested on grade level using Form T at posttest. Test reliability estimates indicate strong subtest internal consistency with coefficient value of .89 for Form T Comprehension (MacGinitie et al.).

**Posttest Procedures**

Posttesting for the SS and TP groups occurred after 18 weeks of instruction at the end of the second quarter. To minimize class disruption, group posttesting was administered in one 90-minute session in the school library. The posttest battery consisted of the following measures: (a) the Cloze Test (3 minutes), (b) the Strategy-Use Test (25 minutes), and (c) the Reading Comprehension section of the GMRT-4 (35 minutes for testing).
Due to a schedule conflict, SSR students were assessed during consecutive days within five days of posttesting of the SS and TP groups. During the first session, the Reading Comprehension section of the GMRT-4 was administered. On the following day, the Cloze Test and the Strategy-Use Test were given. To ensure standardization of testing procedures, the research team, including the authors, proctored all posttesting.

**Treatment Integrity Checklists**

To assess the integrity of the SS teachers' implementation of the specified instruction in each lesson, a treatment integrity (TI) checklist was used. The TI checklist consisted of items classifying instructional practices, routines, and intervention-specific components, such as (a) review of the previous lesson, (b) use of advance and post-organizers, (c) encouraging independent strategy use, (d) giving notes and defining literary terms related to SS instruction, (e) periodically engaging students in dialogue, (f) teacher demonstration and modeling of the SS strategies, and (g) teacher-led practice with SS related strategies. The TI checklist score was calculated by summing the number of observed instructional behaviors for the intervention and dividing it by the total number of points possible, which ranged from 7 to 12 teacher behaviors.

Trained research team members observed and completed checklists for six observations of the SS teachers' instruction. Overall, TI was calculated by summing the number of observed instructional behaviors for the intervention and dividing it by the total number of points possible, for a TI rate of 93% across both SS teachers, ranging from 75% to 100%.

To ensure differences between the instructional conditions, or to assess the extent of content overlap between interventions, procedures for calculating component and average daily treatment integrity were followed (see Gresham, Gansle, Noell, & Cohen, 1993). First, the 29 components of the SS intervention were identified (i.e., 9 associated with self-questioning, 10 each for both story-structure analysis and summary writing). Next, trained research team members observed and completed checklists for four or five visits of the TP and three visits of the SSR teachers' instruction. Finally, observer ratings for each component of the SS intervention and average daily integrity across all components were calculated (see Table 2). Percentages reflect limited overlap between the SS instruction and TP (4 of the 29 components were observed an average of 6.7% to 36.7% of the time, with daily rates of .69% to 2.75%; the remaining 25 com-

<table>
<thead>
<tr>
<th>Table 2</th>
<th><strong>Observer Ratings for Each Component of the Story Structure Intervention and Average Daily Integrity Across All Components</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td><strong>Typical Practice</strong></td>
</tr>
<tr>
<td></td>
<td>RS1</td>
</tr>
<tr>
<td>Self-Questioning</td>
<td>20%</td>
</tr>
<tr>
<td>Identify Main Character</td>
<td>40%</td>
</tr>
<tr>
<td>Identify Central Conflict</td>
<td>0%</td>
</tr>
<tr>
<td>Summary Writing</td>
<td>20%</td>
</tr>
<tr>
<td>25 Remaining Components</td>
<td>0%</td>
</tr>
<tr>
<td>Daily</td>
<td><strong>2.75%</strong></td>
</tr>
</tbody>
</table>

*Note: The percentages reflect the total number of components that TP teachers implemented divided by the total number of components that could have been implemented across all sessions observed and multiplied by 100. The average reflects only the average components observed across the three RS teachers. RS = Reading specialist delivering TP (typical practice) instruction; SSR = Sustained Silent Reading.*
Table 3
Percentage of Explicit Instruction Components Observed by Condition

<table>
<thead>
<tr>
<th align="left">Provide an advance organizer for the lesson by including at least two of the following:</th>
<th>TP</th>
<th>SS</th>
<th>SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td align="left">+ Preview of material, cueing students to focus on particular information, stating lesson objectives, preteaching unknown information</td>
<td>33</td>
<td>83</td>
<td>0</td>
</tr>
</tbody>
</table>

Use of explicit practice

| Provide opportunities for students to respond to questions | 93 | 100 | 0 |
| Provide individual corrective feedback | 21 | 8 | 0 |
| Provide group-oriented guided practice (monitoring student practice carefully, correcting and remediating errors soon after the initial instruction is provided) | 43 | 50 | 0 |
| Provide independent practice (providing opportunities for students to practice by themselves while doing homework or seatwork) | 64 | 17 | 100 |
| Concludes the lesson with a review of what was learned | 0 | 92 | 0 |

Other areas of instruction

| Formative evaluation to measure learning | 0 | 0 | 0 |
| Models (using a think-aloud) the components of a strategy to use when reading, showing student what is expected of them | 7 | 8 | 0 |

Note. TP = Typical Practice; SS = Story Structure; SSR = Sustained Silent Reading.

Interscorer Reliability

To calculate interscorer reliability, two trained scorers independently scored a random sample (20%) of each of the posttest measures, 43% of content checklists, and 23% of pedagogy checklists. Observer recordings were compared using item-by-item analysis. For the Cloze, scorers agreed on 392 out of 393 opportunities to agree (range: 89%-100%) for 99.7% agreement. For the Strategy-Use Test, scorers agreed on 607 out of 608 opportunities to agree (range: 96.8%-100%) for 99.8% agreement. For the GMRT-4, scorers agreed on 770 out of 771 opportunities to agree (range: 97.5%-100%) for 99.9% agreement. For fidelity of treatment across both SS teachers, scorers agreed on 57 out of 60 opportunities to agree (range: 86-100%) for 95% agreement. For content checklists of component and average daily treatment integrity, scorers agreed on 174 out of 174 opportunities for 100% agreement. For pedagogy checklists, scorers agreed on 52 out of 54 opportunities to agree (range: 89-100%) for 96% agreement.

Results

Means and standard deviations for the three posttest measures are presented in Table 4 by grade level and condition. Posttest scores were subjected to three two-way analyses of variance with two levels of grade (fifth, sixth) and three levels of condition (SS, TP, SSR). An alpha level of .05 was used to determine significance of main effect and interactions.

Cloze results suggested no main effect for grade, $F(1,74) = 3.529, p = .064$, or condition, $F(2,74) = 2.174, p = .121$. However, results revealed a significant grade-by-condition interaction, $F(2, 74) = 3.434, p = .037$. Post-hoc confidence intervals were calculated to deter-
### Table 4
**Posttest Outcome Measures**

<table>
<thead>
<tr>
<th>Measure</th>
<th>5th Grade</th>
<th>6th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_1 (SD_1)$</td>
<td>$M_2 (SD^2)$</td>
</tr>
<tr>
<td>Cloze</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR1 vs. SS2</td>
<td>17.62 (2.20)</td>
<td>15.94 (5.64)</td>
</tr>
<tr>
<td>SSR1 vs. TP2</td>
<td>17.62 (2.20)</td>
<td>18.50 (5.39)</td>
</tr>
<tr>
<td>TP1 vs. SS2</td>
<td>15.94 (5.64)</td>
<td>18.50 (5.39)</td>
</tr>
<tr>
<td><strong>Strategy-Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR1 vs. SS2</td>
<td>3.75 (1.39)</td>
<td>3.50 (2.59)</td>
</tr>
<tr>
<td>SSR1 vs. TP2</td>
<td>3.75 (1.39)</td>
<td>2.59 (2.35)</td>
</tr>
<tr>
<td>TP1 vs. SS2</td>
<td>2.59 (2.35)</td>
<td>1.010</td>
</tr>
<tr>
<td><strong>GMRT-4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR1 vs. SS2</td>
<td>21.00 (7.65)</td>
<td>20.83 (6.44)</td>
</tr>
<tr>
<td>SSR1 vs. TP2</td>
<td>21.00 (7.65)</td>
<td>19.82 (7.37)</td>
</tr>
<tr>
<td>TP1 vs. SS2</td>
<td>20.83 (6.44)</td>
<td>7.24</td>
</tr>
</tbody>
</table>

* $p < .0083$. 

**Note:** The table provides a comparison of posttest outcome measures for different conditions, including Cloze, Strategy-Use, and GMRT-4, across 5th and 6th grade levels. The measures include mean scores ($M_1, M_2$), standard deviations ($SD_1, SD^2$), and Cohen’s $d$ effect sizes, which indicate the magnitude of effect. The starred values denote significance at $p < .0083$. The SD Pooled column represents the pooled standard deviation, and $M_1 (SD_1)$ and $M_2 (SD^2)$ denote the mean scores with their standard deviations.
mine if there were differences between conditions (i.e., SS vs. TP, SS vs. SSR, and TP vs. SSR) within each grade level using a two-tailed Bonferroni correction ($\alpha = .025/3$). Results of the 95% confidence interval (95% CI) were not statistically significant between any of the fifth-grade conditions; however, results were statistically significant between the sixth-grade SS mean scores and the SSR group mean scores ($0.536 < x < 12.23$), as well as between the TP mean scores and the SSR group scores (1.5 $\leq x < 12.59$). Thus, both the SS and the TP mean scores were larger than those of the SSR group, but were not statistically different from one another. See Table 4 for effect sizes for groups on all measures.

Strategy-Use results showed no main effect for grade, $F(1,74) = .824, p = .367$, or condition, $F(2,74) = 1.414, p = .250$. The interaction effect was also nonsignificant, $F(2,74) = 2.201, p = .118$. Thus, the Strategy-Use mean scores were equivalent regardless of grade level or instructional condition.

GMRT-4 Comprehension results showed a significant main effect for grade, $F(1,74) = 4.095, p = .047$, indicating that the average comprehension posttest scores were significantly higher for sixth graders ($M = 24.19, SD = 6.231$) than for fifth graders ($M = 20.41, SD = 6.96$) when collapsing across conditions. However, the main effect for condition was nonsignificant, $F(2,74) = .757, p = .473$. In addition, the interaction effect was also nonsignificant, $F(2,74) = 1.485, p = .233$. Thus, GMRT-4 Comprehension mean scores were different between grade levels, but were equivalent regardless of instructional condition.

Student Response to Instruction Trends

Following the above statistical analyses, a separate descriptive analysis was performed to examine individual student response to instruction within conditions for fifth graders (see Table 5). To determine student response, three separate analyses were conducted. First, as students were selected for this study based on their profile as "at-risk" students, response is defined as scoring above the 50th percentile on both of the standardized reading comprehension measures (e.g., The Cloze and GMRT-4 posttests). Such scores indicate a significant response to instruction and result in meeting exit criteria for Tier 2 instruction in this study. Additionally, given the nature of measuring student response to instruction, it is helpful to examine performance on a typical curriculum-based measure (e.g., the Cloze) and a summative measure (e.g., The GMRT-4).

Results indicated that only two of 37 fifth-grade students (5.4%) who took both standardized comprehension measures scored above the 50th percentile on both measures (both students were near the 70th percentile on the GMRT-4 Comprehension posttest). One student was in the SS condition and one student was in the SSR condition (see Table 5 for individual scores and demographic data).

Finally, because it is believed that some students will not respond to instruction, warranting Tier 3 instruction, data were examined for students scoring below the 30th percentile on both standardized measures. Those scores indicated continued "at-risk" status and the need for more intense instruction. Five of the 37 fifth-grade students (13.5%) scored below the 30th percentile on both measures. Collectively, the descriptive analyses showed less than the 75% expected student response to Tier 2 instruction across all conditions of fifth graders (see Coyne, Simonsen, & Faggella-Luby, 2008).

Discussion

This study investigated the effects of three Tier 2 standard treatment conditions of instruction on the reading comprehension performance of at-risk fifth- and sixth-grade urban middle school students. Results of instruction over 18 weeks were mixed, but lend support to the use of RTI instruction in middle schools, and specifically the benefits of explicit Tier 2 instruction across each of the three primary outcome measures. Thus, the study contributes to the research via the following findings.

First, there is evidence to support the nature of instruction in the behaviors associated with successful reading (i.e., SS and TP conditions) as necessary to impact comprehension in addition to continued practice (i.e., SSR). For example, sixth-grade mean scores for students in the SS and TP conditions were higher on all three measures than for students in the SSR condition (see Table 4). Moreover, the Cloze results statistically significantly favored both instructional conditions over SSR, while the remaining two measures demonstrated moderate effect sizes on the strategy-use measure and large effect sizes on the GMRT-4. Additionally, the lack of statistically significant differences between instructional conditions on any of the three measures supports chosen instructional content.

However, fifth-grade data results provide mixed support for the nature of instruction. Specifically, the SSR means on all three measures were either equal to or below means in the SS condition, with small to moderate effect sizes. Such data support the nature of instruction in addition to practice as above. Yet, the SSR means on all three measures were above the TP condition means, again with small to moderate effect sizes. Such a finding at fifth grade requires further examination of the differences in the intervention conditions and may illustrate critical differences between the
<table>
<thead>
<tr>
<th>Student</th>
<th>Condition by teacher</th>
<th>GMRT-4 raw (percentile)</th>
<th>Cloze # correct</th>
<th>Fall 2007 lexile</th>
<th>Winter 2007 lexile</th>
<th>DRP unit score (spring 2006)</th>
<th>DRP unit score (spring 2007)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Free/reduced-price lunch</th>
<th>ELL status</th>
</tr>
</thead>
<tbody>
<tr>
<td>501</td>
<td>SS-1</td>
<td>36 (69)</td>
<td>25</td>
<td>945</td>
<td>1050</td>
<td>37</td>
<td>57</td>
<td>F</td>
<td>Other</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>502</td>
<td>SSR</td>
<td>37 (72)</td>
<td>20</td>
<td>679</td>
<td>820</td>
<td>n/a</td>
<td>n/a</td>
<td>M</td>
<td>Caucasian</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<30 on post Cloze & <30 on post GMRT-4

<table>
<thead>
<tr>
<th>Student</th>
<th>Condition by teacher</th>
<th>GMRT-4 raw (percentile)</th>
<th>Cloze # correct</th>
<th>Fall 2007 lexile</th>
<th>Winter 2007 lexile</th>
<th>DRP unit score (spring 2006)</th>
<th>DRP unit score (spring 2007)</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Free/reduced-price lunch</th>
<th>ELL status</th>
</tr>
</thead>
<tbody>
<tr>
<td>503</td>
<td>TP-3</td>
<td>7</td>
<td>7</td>
<td>752</td>
<td>549</td>
<td>33</td>
<td>51</td>
<td>M</td>
<td>Caucasian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>TP-3</td>
<td>14</td>
<td>9</td>
<td>887</td>
<td>809</td>
<td>35</td>
<td>51</td>
<td>F</td>
<td>Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>505</td>
<td>TP-3</td>
<td>15</td>
<td>7</td>
<td>554</td>
<td>474</td>
<td>35</td>
<td>37</td>
<td>F</td>
<td>Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>SS-1</td>
<td>15</td>
<td>9</td>
<td>950</td>
<td>917</td>
<td>33</td>
<td>44</td>
<td>M</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>TP-4</td>
<td>16</td>
<td>10</td>
<td>857</td>
<td>666</td>
<td>36</td>
<td>50</td>
<td>F</td>
<td>Caucasian</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. n/a = data were unavailable for this student; ✓ = affirmative status within this category. SS = Story Structure; TP = Typical Practice. Unique condition-by-teacher numbers indicate specific teachers within condition (e.g., SS-1 vs. SS-2).
needs of fifth- and sixth-grade struggling readers, potentially related to the explicitness of instruction necessary to improve outcomes.

Second, while the content of instruction associated with successful reading may be established, Tier 2 interventions likely require substantial instructional intensity to improve outcomes for struggling students. In general, data collected during treatment integrity observations specific to instructional pedagogy confirmed that the two instructional conditions failed to consistently implement certain components of direct and explicit instruction as identified by Swanson and colleagues (i.e., Swanson, 1999; Swanson & Deshler, 2003). Specifically, practices such as providing individual and corrective feedback, providing instructional models (via a think-aloud) when introducing a strategy, and using formative assessment to drive instruction were observed in fewer than 21% of instructional sessions (see Table 3).

Limited use of explicit pedagogical practices may be a significant factor in explaining study results. For example, the SS group provides one point of reference, as group means on the strategy-use measure yielded mean scores of 4.67 (grade 5) and 4.43 (grade 6) out of 25 possible points. Such low overall scores are contrary to prior findings by Faggella-Luby and colleagues (2007), in which similar content with more explicit pedagogy was taught (i.e., more modeling, corrective feedback, cooperative learning, independent practice and use of advance/post organizers). Moreover, the poor fifth-grade TP mean scores across measures compared to those of the SSR group may indicate a need for stating lesson objectives, preteaching information to be learned, and reviewing previous instruction and lesson content at the conclusion of the lesson — all areas observed 15 to 92% less in the TP condition than the SS condition.

Such overall limited response (35 of 37 students scoring below the 50th percentile on both standardized measures after 18 weeks of instruction), along with five students continuing to score below the 30th percentile on both standardized measures, may be a direct result of the lack of explicit instruction that accompanied both SS and TP. Moreover, though instruction continued for 18 weeks, the limited time in instruction (2-3 days per week for 30 minutes) resulted in mean instructional times ranging from 16 to 22 hours total. Consequently, students receiving Tier 2 instruction may require a more intense dosage of explicit instruction over fewer total weeks to achieve adequate response. Second, limited response on the Cloze (especially by SS students) may be attributed to the lack of instructional focus on sentence-level comprehension. More specifically, the intent of instruction was to improve passage-level comprehension, with possible limited application for the Cloze task.

Third, story-structure knowledge remained limited for struggling fifth- and sixth-grade students, necessitating further investigation as a standard treatment intervention. Data from the strategy-use measure across groups clearly indicate that story-structure knowledge was not part of the prior knowledge of struggling readers. In fact, even after instruction, students had not achieved mastery of this material. Specifically, the posttest means for all groups were below 19%, indicating substantial difficulties with identifying story-structure components, self-questioning, and summary writing. As indicated above, further investigation is necessary to determine the level of instructional intensity necessary for student mastery of these seemingly advanced but critical reading comprehension skills.

Fourth, results showed that it is not too late to provide intervention in the middle grades for struggling readers. The significant main effect for grade level on the GMRT-4 Comprehension measure indicates that although students in both instructional conditions were low performing, their comprehension abilities had not plateaued. This finding again supports the use of RTI as a framework for delivering increasingly intensive instruction to improve student performance. Additionally, the lack of significance between groups on the GMRT-4 posttest mirrors closely the trends on the Cloze, in which sixth-grade SS and TP mean scores were both approximately five raw score points above SSR (resulting in large effect sizes for both comparisons; see Table 4), again indicating support for Tier 2 instruction and the RTI framework. The fifth-grade trend was also similar to the previous findings, in which SS and SSR means were higher than TP means.

Finally, the lack of statistical difference between instructional conditions may be of significance to high-poverty middle schools where teacher turnover is higher than in low-poverty schools (Planty et al., 2008). This study compared two first-year teachers (bachelor degrees only) delivering the SS instruction with three certified and experienced reading specialists. High turnover rates in urban districts often result in early-career teachers working with struggling students in conditions that may parallel those in the current study (Andrews & Quinn, 2004). Therefore, a finding of non-significance between groups may mean that the SS intervention provided the necessary supports for early-career teachers to help their students make gains similar to those when compared to veteran peers.

**Implications for Practice**

As previously discussed, Fuchs and Deshler (2007) posed several questions related to applications of the
RTI framework in middle school that are addressed in this study. First, RTI is very relevant at the middle school level. Comparison of student response to instruction, especially the SS condition, to an SSR group indicates support for Tier 2 instruction in addition to reading practice. Demographic and posttest data from this study underscore the extraordinary need for an instructional framework that will close the achievement gap between struggling middle school learners and their typically achieving peers through supplemental explicit instruction.

Second, a key attribute of the multitier framework is the provision of increasingly intense service delivery beyond changes in the size of instructional groups. This study provides empirical evidence to support the need for instruction in what good readers do, while underscoring the urgent need for appropriate use of explicit pedagogical practices at Tier 2. Third, Tier 2 instruction was conducted by leveraging the malleable components of school resources. Initially, TP condition reading specialists were reassigned from individual student meetings to small-group instruction for 2 to 3 days per week. Additionally, through a university-school partnership, two interns were made available to build the capacity of the program and increase the number of students who could receive additional instruction. Missing from this configuration was an onsite coordinator to manage schoolwide testing, data collection, and fidelity assessment.

Finally, TI assessment, in particular, requires (a) school personnel (preferably a building-level administrator) to coordinate and observe, (b) measures that focus on instructional intensity and explicitness, (c) regular checks of core/universal instruction to ensure implementation, (d) clear communication between teachers and observers on the purpose of the assessment, and (e) regular and timely feedback based on observations.

**Limitations**

Results from this study must be considered in light of several study limitations. First, although randomization of at-risk subjects with posttest-only dependent measures is considered a rigorous research design (McMillan, 2008), it does not provide a complete profile of student performance consistent with RTI. For example, due to contextual factors beyond researcher control within the middle school, regular progress monitoring was impossible to conduct with reliability. Moreover, it was impossible for the school to secure current high-stakes assessment data during the spring semester to facilitate accurate student placement in supplemental instruction for the following school year (resulting in the use of the previous year’s data).

Second, small sample size limits the power in our statistical comparisons. However, within this study, school resources were exhausted (3 teachers providing additional Tier 2 instruction), necessitating the two interns to serve additional students within the subject population. Consequently, replication of the study across multiple school sites is necessary to increase the sample size while realistically achieving acceptable student-to-teacher ratios within a school’s limited capacity.

Third, while the study was intended to test the RTI framework by measuring student response to instruction at Tier 2, overall conditions to support successful implementation described by Fuchs and Deshler (2007) were not implemented consistently. Specifically, implementation lacked (a) significant and sustained investments in professional development, (b) district-level support, (c) teachers who were willing to embrace new roles, and (d) practitioner involvement in the decision to adopt this model. Regrettably, failure to implement many of these features was the direct result of a sudden administrator health emergency, underscoring the importance of constructing a building-level team during adoption consideration. In addition, given the reality of applied research in school settings (e.g., snow days and bus vandalism causing school cancelation; false fire alarms taking limited class time), the extended instructional time (from past studies) was not as intense as anticipated.

Fourth, students with reading disabilities were excluded from the study at the request of the school, limiting generalization of findings to students with LD. Finally, overall limited responsiveness to instruction, especially on the strategy-use measure, may indicate the need for more explicit pedagogical practices with struggling middle school learners to maximize the impact of the limited overall time-in-instruction. Although overt and explicit instruction provides the most accurate magnitude of treatment outcomes (Swanson, 1999), explicit instruction was not consistently observed across fidelity measures in either instruction condition.

**Future Research**

Future research on both Tier 2 reading comprehension interventions, in general, and a Story-Structure Routine specifically, should attempt to answer the following questions:

- What is the intensity or dosage of instruction necessary to achieve significantly greater numbers of students who respond to instruction by moving into the > 50 percentile range on standardized assessments?
- How might regular feedback to teachers from fidelity measures improve student outcomes?
• What assessments are currently being given that can be used to inform instructional decisions (e.g., core reading tests, curriculum-based measures like the Cloze, universal screening)?
• What level of professional development in explicit pedagogy would be necessary to provide greater instructional intensity?
• What are the malleable components of the middle school context, and how can researchers and practitioners continue to collaborate effectively and efficiently to learn more about design and implementation of feasible evidence-based practices?

In summary, while much has been learned, considerably more technical information is necessary to ensure RTI as a viable model for improving student outcomes in middle school, with considerable attention beyond the content of instruction (what reading comprehension is taught), toward the pedagogy of instruction (how reading comprehension is taught). Specifically, models for Tier 2 instruction require closer examination of their instructional dosage, including total time in instruction, impact of pedagogy on variations in content of instruction, and levels of explicitness in the delivery of instruction. Practitioners and researchers are called to work together to find solutions that will address the challenges of effectively educating struggling middle school learners.

References
Simonsen, B., Shaw, S., Faggella-Luby, M., Sugai, G., Coyne, M.,


**Author Notes**

Examination of sixth-grade students was not possible, as percentiles for the Cloze measure were based on the fifth-grade passage and, therefore, would be inflated. Additional student data are in the form of school-collected measures to further describe the response subject population. Of note is the significant degree of variability of student performance on these measures.

Please address correspondence about this article to: Michael Faggella-Luby, University of Connecticut, Center for Behavioral Education and Research, 249 Glen Brook Road, Unit 2064, Storrs, CT 06269-2064; e-mail: mike.fl@uconn.edu
Copyright of Learning Disability Quarterly is the property of Council for Learning Disabilities and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.