This study examined the effects of story mapping plus incentives on writing proficiency. A single-subject multiple baseline design was used in which 16 students (1 fifth, 1 fourth, 14 second, and 1 first grade) were provided direct instruction in story mapping. Incentives were used as reinforcers for students who wrote 30 percent more words during intervention than they did during baseline. Total written words, words spelled correctly, correct punctuation marks, correct word sequences, percentage of words spelled correctly, percentage of correct word sequences, and correct minus incorrect word sequences were used to assess written expression skills. Visual inspection was used to examine data for changes in level and/or slope.
THE EFFECTS OF STORY MAPPING AND INCENTIVES
ON MULTIPLE MEASURES OF WRITING PROFICIENCY

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Specialist in Education
by
Melissa Ann Brunner
Miami University
Oxford, Ohio
2010

Advisor: ____________________
Dr. T. Steuart Watson

Reader: ____________________
Dr. Kevin Jones

Reader: ____________________
Dr. William Boone

Reader: ____________________
Dr. Michael Woodin
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>Method</td>
<td>12</td>
</tr>
<tr>
<td>Results</td>
<td>15</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>References</td>
<td>29</td>
</tr>
<tr>
<td>Appendices</td>
<td>33</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Production Dependent Indices 4
Table 2: Production Independent and Accurate Production Indices 6
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total Written Words</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Words Spelled Correctly</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Correct Word Sequences</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>Correct Punctuation Marks</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>Percentage of Words Spelled Correctly</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>Percentage of Correct Word Sequences</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>Correct Minus Incorrect Word Sequences</td>
<td>24</td>
</tr>
</tbody>
</table>
LITERATURE REVIEW

Because written expression is necessary for success in most academic content areas, it is crucial that students receive adequate instruction in the components encompassed by proficient writers (Gansle, VanDerHeyden, Noell, Restar, & Williams, 2006). As such, reliable methods that formatively evaluate student writing are essential to providing instruction because they allow educators to identify strengths and weaknesses and modify and/or develop interventions when weaknesses are found. Objectively assessing written expression can prove difficult due to the complexity of the task and the number of components necessary for good compositions (Isaacson, 1988).

Currently, written expression is measured using such devices as percentage of story elements, holistic teacher judgments, standardized assessments, and curriculum based measurement (CBM) (Espin, et al, 1999; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002; Lane, Harris, Graham, Brindle, & Morphy, 2008). Percentage of story elements is calculated by summing the number of story elements in a written work and dividing it by the number of possible story elements. Although the terminology differs, story elements usually encompass variations of the following: setting, time, characters, problem, solution, outcome, reaction, and theme (Albertson & Billingsley, 2001; Albertson, 1998; Lane, Harris, Graham, Brindle, & Morphy, 2008). Although frequently used, percentage of story elements, does not assess for writing conventions such as the grammar, structure, and organization needed to write a cohesive, meaningful story.

Apart from measuring writing using strictly quantitative measures, holistic teacher judgments are assessments based upon an established rubric that provides the components necessary for a written passage at different levels of proficiency. This criterion usually includes measures of syntactic and semantic maturity, content, and conventions (Howell & Nowlet, 2000). Despite their widespread uses, holistic teacher judgments are often subject to rater bias, which is demonstrated through differences in scoring based on handwriting, punctuation, and spelling. In fact, many researchers correct passages for these errors before scoring (Regan, Mastropieri, & Scruggs, 2005; De La Paz, 1997; Reid & Lienemann, 2006; Lane et al., 2008).

Although standardized assessments are excellent at determining a student’s present level of performance, they are not useful for the data-based decision making needed to plan instruction because they typically do not assess skills using tasks similar to those required in academic
settings (Isaacson, 1988). Thus, they have little face validity when designing interventions. Additionally, Deno, Mirkin, and Marston (1980) reported that standardized tests are too time consuming and expensive to use repeatedly. They are also insensitive to change due to the small number of item sets that assess individual skill, and the stability of scores on standardized assessments makes them poor measures of short-term growth (Gansle et al., 2004). In contrast, Deno, et al. (1980) stated that formative evaluation measures must incorporate several components in order to be successful for their intended purpose. These components include: validity in relation to other assessment measures, sensitivity to intervention, and ease of administration. They must also be inexpensive, time efficient, unobtrusive, and easy to use repeatedly. While standardized assessments are good for summative evaluations, CBM are better at formatively evaluating student progress in a Response to Intervention (RtI) framework (Deno, et al., 1980).

**CBM: Valid Indices of Written Expression**

A substantial portion of the literature on formative evaluation shows that CBM encompasses all of these fundamentals in direct relation to the curriculum (Jenkins, Deno, & Mirkin, 1979; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). More specifically, CBM indices differentiate between students of differing abilities (Tindal & Parker, 1989; Watkinson & Lee, 1992), and are valid in relation to standardized assessments (Deno, et al., 1980; Espin, Scierka, Skare, & Halverson, 1999; Espin et al., 2000; Gansle et al., 2002; Gansle et al., 2004; Jewell & Malecki, 2005; Tindal & Parker, 1989; See Table 1).

Additionally, CBM indices are useful for measuring different types of skills, including a student’s fluency and accuracy with writing. Production dependent indices measure fluency, meaning that a student’s score is based on how much he or she produces or writes within a specific period of time. In contrast, production independent indices assess a student’s accuracy when writing regardless of how much the student writes. Finally, accurate production indices measure both fluency and accuracy (Espin et al., 2000; Jewell & Malecki, 2003). These indices may be evaluated individually or in concert with other indices, depending on the type of writing skills assessed.

**Production Dependent Indicies**

**Total Written Words (TWW):** TWW is a production dependent measure that is calculated by tallying the total number of written words regardless of spelling, capitalization, and grammar
Deno (1981) investigated the reliability of TWW, including: comparison over time, comparability of forms, internal consistency, and administration and scoring and found moderate to strong correlations. Additionally, Gansle et al. (2004) found that TWW is sensitive to intervention effects over a short period of time.

The following table depicts the correlations between TWW and other production dependent indices with many criterion related variables, including both standardized assessments and holistic ratings:
Table 1: Production Dependent Indices

<table>
<thead>
<tr>
<th>Correlation with TWW</th>
<th>Correlation with WSC</th>
<th>Correlation with CPM</th>
<th>Correlation with CWS</th>
<th>Criterion Related Variable</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>r = .05 to .17</td>
<td>r = .08 to .21</td>
<td>r = .18 to .31</td>
<td></td>
<td>California Achievement Test (Language Subtests)</td>
<td>Espin et al., 1999</td>
</tr>
<tr>
<td>r = .84</td>
<td></td>
<td></td>
<td></td>
<td>Developmental Sentence Scoring</td>
<td>Deno et al., 1982</td>
</tr>
<tr>
<td>r = .43 to .47</td>
<td>r = .46 to .51</td>
<td></td>
<td></td>
<td>District Writing Test</td>
<td>Espin et al., 2000</td>
</tr>
<tr>
<td>r = .22 to .25</td>
<td>r = .25 to .29</td>
<td></td>
<td></td>
<td>English Grade Point Average</td>
<td>Espin et al., 1999</td>
</tr>
<tr>
<td>r = .36</td>
<td>r = .41</td>
<td></td>
<td></td>
<td>Holistic Ratings (5 Point Scale)</td>
<td>Espin et al., 1999</td>
</tr>
<tr>
<td>r = .34 to .46</td>
<td>r = .38 to .48</td>
<td></td>
<td></td>
<td>Holistic Ratings (4 Point Scale)</td>
<td>Espin et al., 2000</td>
</tr>
<tr>
<td>r = .10</td>
<td>r = .31</td>
<td></td>
<td></td>
<td>Holistic Ratings</td>
<td>Tindal &amp; Parker, 1989</td>
</tr>
<tr>
<td>r = .10 to .15</td>
<td>r = .18 to .24</td>
<td>r = .36 to .44</td>
<td>r = .36 to .43</td>
<td>Iowa Test of Basic Skills (Writing Subscales)</td>
<td>Gansle et al., 2002</td>
</tr>
<tr>
<td>r = .12 to .45</td>
<td>r = .20 to .51</td>
<td></td>
<td></td>
<td>Language Arts Grade</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
<tr>
<td>r = .16 to .28</td>
<td>r = .26 to .29</td>
<td>r = .25 to .26</td>
<td>r = .28 to .41</td>
<td>Louisiana Educational Assessment Program (Writing Subtests)</td>
<td>Gansle et al., 2002</td>
</tr>
<tr>
<td>r = -.14 to .24</td>
<td>r = -.05 to .38</td>
<td></td>
<td></td>
<td>Stanford Achievement Test (Language and Spelling Subtests)</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
<tr>
<td>r = .62</td>
<td>r = .67</td>
<td></td>
<td></td>
<td>Stanford Word Use</td>
<td>Deno et al., 1982</td>
</tr>
<tr>
<td>r = .08</td>
<td>r = .21</td>
<td>r = .37</td>
<td>r = .36</td>
<td>Teacher Ranking</td>
<td>Gansle et al., 2002</td>
</tr>
<tr>
<td>r = .16 to .44</td>
<td>r = .24 to .49</td>
<td></td>
<td></td>
<td>Tindal and Hasbrouck Analytic Scoring System</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
<tr>
<td>r = .58 to .75</td>
<td>r = .57 to .80</td>
<td></td>
<td></td>
<td>Test of Written Language</td>
<td>Deno et al., 1982</td>
</tr>
<tr>
<td>r = .23</td>
<td></td>
<td></td>
<td></td>
<td>Woodcock Johnson-Revised (Writing Samples Subtest)</td>
<td>Gansle et al., 2004</td>
</tr>
</tbody>
</table>
Words Spelled Correctly (WSC): WSC is another production dependent index that is the summation of the number of words that are spelled correctly in a written work. Words spelled correctly, but incorrectly within the context of the written passage, are counted as correctly spelled (Gansle et al., 2002). Deno (1981) found that test-retest reliability had the lowest correlations (one day: \( r = .57 \) to \( r = .92 \), three weeks: \( r = .50 \) to \( r = .70 \)) while the reliability of forms, internal consistency, and administration and scoring had correlations ranging from \( r = .70 \) to \( r = .98 \). WSC did not distinguish between students of different ability levels (Tindal & Parker, 1989; Watkinson & Lee, 1992).

Correct Punctuation Marks (CPM): Another production dependent index is CPM, which is found by evaluating all the punctuation marks in a writing sample to determine if each mark is applied correctly within the sentence. The number of correct marks comprises the CPM score. Omitted punctuation does not affect CPM (Gansle et al., 2002). Gansle et al. (2002) examined CPM in relation to several criterion related variables and found weak correlations.

Correct Word Sequences (CWS): CWS is also a production dependent measure of students’ writing fluency. CWS is calculated by determining the total number of sequences or two adjacent, correctly spelled words that are acceptable within the context of the phrase by someone whose first language is English (Videen, Deno, & Marston, 1982). CWS takes into account writing conventions such as grammar, major punctuation, and spelling (Jewell & Malecki, 2005; Watkinson & Lee, 1992).

Watkinson and Lee (1992) found statistically significant differences between students with LD and students without LD on CWS. Conversely, Tindal and Parker (1989) found that CWS did not differentiate between groups of students. However, Tindal and Parker’s study compared students with LD to students in remedial programs. Thus, it is logical to conclude that there are bigger differences between students with and without LD than there are between students with LD and those in remedial programs.

Production Independent Indices

Percentage of Words Spelled Correctly (%WSC): %WSC is a production independent variable. This means that it is an indicator of students’ accuracy during writing. %WSC is calculated by dividing WSC by TWW and multiplying by 100 (Tindal & Parker, 1989). For example, students who write 30 WSC out of 60 TWW would score 50 percent on %WSC. This is contrasted with students who write 30 WSC out of 30 TWW and score 100 percent. Tindal
and Parker (1989) showed that %WSC showed significant differences between special education and remedial groups of students while WSC did not.

The following table depicts the correlations between %WSC, other production independent and combined indices with many criterion related variables:

Table 2: Production Independent and Accurate Production Indices

<table>
<thead>
<tr>
<th>Correlation with %WSC</th>
<th>Correlation with %CWS</th>
<th>Correlation with CMIWS</th>
<th>Criterion Related Variable</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>District Writing Test</td>
<td>Espin et al., 2000</td>
</tr>
<tr>
<td>r = .69 to .75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Holistic Ratings (4 Point Scale)</td>
<td>Espin et al., 2000</td>
</tr>
<tr>
<td>r = .65 to .70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = .73</td>
<td>r = .75</td>
<td></td>
<td>Holistic Ratings</td>
<td>Tindal &amp; Parker, 1989</td>
</tr>
<tr>
<td>r = .45 to .53</td>
<td>r = .29 to .58</td>
<td>r = .36 to .61</td>
<td>Language Arts Grade</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
<tr>
<td>r = .43 to .52</td>
<td>r = .44 to .67</td>
<td>r = .41 to .62</td>
<td>Stanford Achievement Test (Language and Spelling Subtests)</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
<tr>
<td>r = .34 to .39</td>
<td>r = .40 to .49</td>
<td>r = .54 to .56</td>
<td>Tindal and Hasbrouck Analytic Scoring System</td>
<td>Jewell &amp; Malecki, 2005</td>
</tr>
</tbody>
</table>

**Percentage of Correct Word Sequences (%CWS):** %CWS is another important production independent CBM measure. It is found by calculating the number of correct word sequences divided by the total number of words sequences and multiplying by 100 (Tindal & Parker, 1989). As such, it is a production independent indicator because the quality of work is assessed rather than the amount of work produced. For example, using the same principles as the %WSC example, a student who writes 50 CWS out of 100 total word sequences achieves a lower score than a student who writes 15 CWS out of 20 total word sequences.

%CWS was the best measure examined by Tindal and Parker (1989) at distinguishing between students in special and remedial education programs. Additionally, Watkinson and Lee (1992) showed that %CWS differentiated between students with and without disabilities. However, these same studies found that CWS alone did not distinguish between the same groups.

**Accurate Production Index**

**Correct Minus Incorrect Word Sequences (CMIWS):** CMIWS combines production
dependent and independent indices to provide a measure that evaluates both students’ accuracy and fluency. CMIWS is found by calculating the number of CWS and then subtracting the number of incorrect word sequences (IWS) (Jewell & Malecki, 2005). Espin et al. (2000) and Jewell and Malecki (2005) investigated CMIWS in relation to several criterion related variables and found weak to moderate correlations.

Educators must be aware of the different skills that are assessed by each CBM index. When measures of fluency are needed, production dependent indices such as TWW, WSC, CPM, and CWS should be considered. When accuracy needs to be assessed, production independent variables such as %WSC and %CWS should be used. However, it is important to examine both production dependent and independent indices because the meanings of scores may change when the relationship between scores is considered (ie. 30 WSC/60 TWW = 50% %WSC and 30 WSC/30 TWW = 100% %WSC). Furthermore, correlations to criterion variables differ in strength depending on the index examined, and some indices are better at differentiating between groups of students.

**Strategies That Impact Written Expression Skills**

*Modeling.* Regan et al. (2005) examined the effects of dialogue journals on the compositions of students with emotional disturbance (n = 5). During the multiple baseline intervention phase, the teacher responded to entries, modeled appropriate writing conventions and provided prompts and suggestions. The students increased from an average of 29 to 73 percent time on-task from baseline to intervention, and this continued after the intervention was withdrawn. Additionally, four of the five students increased from an average of 18.5 to 43.3 TWW during intervention, maintaining an average of 39.0 TWW at follow-up. Although there was great variability on quality ratings, four of the students showed statistically significant gains. However, these gains decreased during the maintenance period (Regan et al., 2005).

*Direct Instruction.* Walker, Shippen, Alberto, Houchins, and Cihak (2005) studied the impact of *Expressive Writing* on the performance of three high school students with learning disabilities (LD). *Expressive Writing* is a direct instruction program that teaches editing and the mechanics of writing sentences, paragraphs, and stories. The students increased in CWS from baseline (mean = 22) to intervention (mean = 31) and from intervention to follow-up (mean = 34). They also improved on the Test of Written Language- Third Edition (TOWL-3) from the pre- to post-test.
**Strategy Instruction and Self-Regulation.** Albertson and Billingsley (2001) evaluated the impact of strategy instruction and self-regulation on the written performance of two gifted, seventh grade students who hurried through writing tasks, using a multiple-baseline, time-series design. The students used story mapping and prompts during the intervention phase. They also set goals, charted progress, self-monitored, and received praise when they met their goals. The students met or exceeded 88 percent of their goals by writing more in a shorter time and including more story elements. They also improved on holistic quality.

**Self-Regulated Strategy Development.** Graham and Harris (1993) developed Self-Regulated Strategy Development (SRSD) and examined its effectiveness with positive results during several group studies (Graham & Harris, 1993; 1999). SRSD is a seven stage program that includes development of needed pre-skills, discussion of baseline and intervention, modeling, memorization of steps, assisted practice, and independent practice of skills taught to a specific criterion. Additionally, self-regulation techniques including goal setting, self-monitoring, and self-reinforcement are taught and faded.

De La Paz (1997) examined SRSD on the opinion essays of three fifth grade students with LD. A multiple probe design was used to teach the students advanced planning to mastery. Papers were corrected for spelling, punctuation, and capitalization prior to scoring. Overall, the students spent less than six minutes writing and almost no time planning during baseline, which resulted in short essays of poor quality. Conversely, after SRSD instruction, the students’ essays were two to four times as long, included more functional elements, and were more coherent. These gains were mostly maintained at follow-up.

Lienemann, Graham, Leader-Janssen, and Reid (2006) also studied the SRSD approach with six second grade students with and without disabilities. During intervention, direct instruction focused on identifying story elements while reading, modeling of writing the story elements, planning with assistance, and independent practice. The students improved from baseline to intervention on TWW, number of story elements, and holistic ratings. Improvements in TWW and story elements were maintained at follow-up. However, holistic ratings decreased but not to baseline levels. Generalization was assessed by the students’ ability to identify story elements when reading. Four children recalled more story elements during intervention than baseline.

Reid and Lienemann (2006) investigated the effect of SRSD on the written performance
of three students with Attention Deficit Hyperactivity Disorder (ADHD). Stories were edited for spelling, punctuation, and capitalization errors prior to measurement. The students were taught writing and planning strategies. The mean number of story elements and TWW improved from baseline to intervention and were maintained at follow-up. Quality ratings also improved after intervention, but decreased from intervention to follow-up. However, they were still above baseline levels.

Lane et al. (2008) studied SRSD with six second-grade students who were at risk for behavior problems. A school-wide positive behavior support program, in which students received tickets for behaviors reflecting the core values of the school was implemented throughout the intervention. The students could earn one ticket per session. Papers were corrected for spelling, punctuation, and capitalization prior to scoring. Only one student planned during baseline while all of the students planned during intervention and maintenance. Following intervention, story length, quality ratings, and the number of story elements increased. Results were generally maintained. As the intervention was administered following a multiple baseline across participants design, the maintenance phase varied from 2 to 11 weeks (Lane et al., 2008). Although the effectiveness of the SRSD approach is undeniable, the sheer number of components that it includes limits the ability to assess which components are the most effective at improving written expression skills.

**Story Mapping.** Vallecorsa and deBettencourt (1997) examined the effects of direct instruction of story mapping elements on the reading and writing skills of three students with LD. After baseline was established, the students received explicit instruction in reading that included teacher modeling, verbal prompting, direct questioning, and independent practice using story elements. Although the students improved in measures of reading comprehension, their knowledge of the story elements did not affect their written performance. Thus, explicit instruction using the same elements to plan for written expression tasks was implemented and explicit instruction in reading was discontinued. During this phase, one student improved in writing while the other two improved in both reading comprehension and writing.

In contrast to these findings, Idol (2001) studied the reading performance of 27 third and fourth grade students of varying abilities. The students were provided story-mapping instruction. Reading comprehension scores increased from baseline for all students with low achieving students showing similar gains as higher achieving students. Additionally, a post-hoc procedure
found that students’ writing in their daily journals showed increases in the inclusion of story elements from the beginning to the end of the study although writing was not specifically targeted. The student in the story mapping condition increased by writing an average of three additional story elements while the students in the control group only increased by an average of two. Moreover, the low achieving students showed the greatest improvement with a statistically significant increase of 3.6 written story elements from the beginning to the end of the study.

Sturm (1996) compared the effects of hand-drawn and computer-generated concept maps on the writing performance of 24 eighth grade students with disabilities. Significant improvements were demonstrated during both mapping conditions for students with LD on TWW, number of T-units, and holistic writing. Significant improvements were also demonstrated by the students with reading disabilities on holistic writing scores. The outcome effects of hand-drawn and computer-generated concept maps were similar across groups and measures. However, the students with LD self-reported that they enjoyed using the computer-generated maps more than the hand-drawn maps. Carryover effects limited the overall conclusiveness of the findings from this study.

Incentives. Hosie, Gentile, and Carroll (1974) investigated the effects of the Premack Principle with 28 sixth grade students. Hosie et al. observed the students during a free period. After their preferred activity was determined, the students were provided an opportunity to read books and complete written book reports in order to access the free time materials. The students were randomly assigned into conditions so that only half of them received their preferred activity contingent on the completion of the written report. Eleven students in the preferred contingency group completed the written assignment while only two students in the non-preferred contingency group completed it. Moreover, the preferred contingency group completed it faster.

Dickerson and Creedon (1981) compared the effects of self- versus teacher-selected standards on the written and math performance of 30 second grade and third grade students. The students were divided into three equal groups that were matched for grade, ability, and gender and included pupil-selected, teacher-selected, and no contingencies. However, the pupil- and teacher-selected groups were yoked so that the pupil-selected group determined their standards as well as those of the teacher-selected group. All of the students were given ten minutes to complete a writing assignment that was scored for TWW and another ten minutes to complete a math assignment that was scored for the number of problems correct. There were no statistically
significant differences across groups during the baseline conditions.

Both the pupil- and teacher-selected groups increased significantly on TWW; however, the pupil-selected groups showed a much greater increase than the teacher-selected group. The no contingency group showed no significant change. The pupil-selected group also increased significantly on the number of correct math problems while the teacher-selected group improved but not significantly and the control group showed no significant change (Dickerson & Creedon, 1981). Thus, it appears that students are more willing to complete tasks when they participated in setting the accompanying performance standards.

Bording (1984) studied nine students in special education and the effects of free time on the students’ grammar. Using a multiple baseline design, the students were entered into the free-time contingency in which they were able to earn points that could be exchanged for free time if they improved in capitalization, punctuation, overall accuracy, and neatness. The students’ writing was rated on measures of capitalization, punctuation, subject-verb agreement, overall grammar, cursive writing, and the quality of composition. The students showed improvement on all ratings in the intervention phase. The students with LD responded the most favorably to the intervention, while the students who were behaviorally disabled showed some improvement, and the students with cognitive disabilities responded the least.

Limitations

Although students respond positively to several types of written expression interventions, direct instruction in story mapping has the most research support (Sturm, 1996; Vallecorsa & deBettencourt, 1997; Idol, 2001). Additionally, there is some evidence that adding incentives to story mapping intervention might strengthen the improvements of written expression skills (Hosie et al., 1974; Dickerson & Creedon, 1981; Bording, 1984; Lane et al., 2008). Finally, CBM has been validated as a means for problem identification through numerous empirical studies (Deno et al., 1980; & Parker, 1989; Watkinson & Lee, 1992; Espin et al., 1999; Gansle et al., 2002; Gansle et al., 2004; & Malecki, 2005). However, there is not a consensus amongst educators on which CBM indices are best for progress monitoring within an RTI model of service delivery.

Therefore, this study examined the effects of story mapping plus incentives on writing proficiency. A single-subject multiple baseline design was used in which 16 students (1 fifth, 1 fourth, 14 second, and 1 first grade) were provided direct instruction in story mapping.
Incentives were used as reinforcers for students who wrote 30 percent more words during intervention than they did during baseline. TWW, WSC, CPM, CWS, %WSC, %CWS, and CMIWS were used to assess written expression skills. Visual inspection was used to examine data for changes in level and/or slope.

**METHOD**

*Participants/Setting*

The participants in this study were 16 (male = 13 and female = 3) students from a rural area of Southwestern Ohio. Thirteen of the students were entering third grade while the other three students were entering second, fourth, and fifth grades respectively. The students were from low- to middle-class socioeconomic status. They were recommended by their school district to participate in a summer academic program due to their high risk for poor reading outcomes. The students participated in either 2006 or 2008. The 12 (2008) or 14 (2006) day long program was run by first and second year school psychology graduate students under the supervision of school psychology faculty and took place in a typical classroom setting.

*Procedures*

**Baseline:** To establish baseline, the students were given a story starter and instructed to draw a picture for five minutes about their story. They were then given five minutes to write. See Appendix A for a list of the story starters and Appendix B for the baseline script. The students individually completed one five-minute timed writing probe daily until a stable trend was determined through visual inspection.

**Treatment:** The students were given a story starter and a story map. They then had five minutes to plan their stories with the assistance of their teachers who asked them questions in order to help them complete their story map. Then the students had five minutes to write their stories using the story map. The students earned a reinforcer if they wrote 30 percent more words during treatment than they did during baseline. See Appendix C for the treatment script.

Story starters were counterbalanced across participants and stages.

*Measurement*

All five-minute timed written passages were scored for the following CBM measures:

**Total Written Words (TWW):** TWW is the total number of written words regardless of spelling, capitalization, and grammar. Units of letters separated by a space are considered a word. Numbers and symbols that are not spelled out are not included. Titles and words copied
from the story starters are included.

*Words Spelled Correctly (WSC):* WSC is the total number of words that are spelled correctly *within* the passage context. For example, “I have went their before.” would count as 4 WSC because although “their” is spelled correctly, it is incorrect within the context of the sentence.

*Correct Punctuation Marks (CPM):* CPM is the total number of correctly applied punctuation marks. Missing marks do not affect this score.

*Correct Word Sequences (CWS):* CWS is the total number of sequences or two adjacent, correctly spelled words that are acceptable within the context of the phrase by someone whose first language is English. Capitalization and major punctuation (.,!?) is included. For example, the sentence “Her go to the stor at 4 o’clock yesterday.” would be scored:

\[ \text{\textasciitilde Her \textasciitilde go \textasciitilde to \textasciitilde the \textasciitilde stor \textasciitilde at \textasciitilde 4 \textasciitilde o'\textasciitilde clock \textasciitilde yesterday \textasciitilde}. \]

The CWS for this sentence is four. “Her” shows a capitalization at the beginning of the sentence and is therefore, counted as a CWS. “Her go” is not syntactically correct and is scored as an incorrect word sequence (IWS). However, “go to” and “to the” are acceptable two word sequences. Since “stor” is spelled incorrectly, both “the stor” and “stor at” are IWS. “at 4,” “4 o’clock,” and “ o’clock yesterday” are CWS. “yesterday.” is counted as an additional CWS because the period is a correctly used major punctuation mark.

*Percentage of Words Spelled Correctly (%WSC):* %WSC is calculated by dividing WSC by the TWW and multiplying by 100.

*Percentage of Correct Word Sequences (%CWS):* %CWS is calculated by dividing CWS by the total number of word sequences and multiplying by 100.

*Correct Minus Incorrect Word Sequences (CMIWS):* CMIWS is calculated by subtracting the number of IWS from the number of CWS.

**Reliability**

First and second year school psychology graduate students administered the intervention under the supervision of the thesis advisor. All examiners completed a graduate level class on the administration and interpretation of educational assessments and had experience administering assessments to children prior to the study. The examiners were trained and reliability checked in scoring the CBM indices. Additionally, all examiners were provided with a scoring guide that explained how to calculate the CBM indices (see Appendix F).
Treatment Integrity

Interrater agreement was attained for 33 percent of the writing probes. Additionally, all teachers completed a scoring rubric following each session to ensure that the students completed the drawing and story maps according to the intervention script (see Appendix E).

Design and Procedures

This study used a multiple baseline across subjects design (Cooper, Heron, & Heward, 1987) to show a causal relationship between story mapping plus incentives and TWW, WSC, CPM, CWS, %WSC, %CWS, and CMIWS. Baseline started at the same time for all participants with treatment beginning after variable lengths of baseline. In order to rule out the effects of testing and measurement, a stable baseline in TWW was attained. Then treatment began for the Three Point Cohort. Treatment for the Five and Seven Point Cohorts began in a stagger fashion after the effects of treatment were established for the preceding cohort. Each child received treatment for three to eleven sessions depending on when their cohort began treatment.

Data Analysis

To determine the effects of story mapping plus incentives on writing proficiency, changes in TWW during baseline were compared to changes in TWW during treatment and following the introduction of story mapping and incentives. If these changes occurred when, and only when, story mapping and incentives were introduced and this effect was replicated across all 16 participants, then one can reasonably conclude that the changes were due to story mapping and incentives. Similar analyses occurred for WSC, CPM, CWS, %WSC, %CWS, and CMIWS.

Protection of Human Subjects

This study was reviewed and approved by an institutional review board (IRB) due to the use of human participants. There was no harm or significant risk to any of the subjects throughout the study. Confidentiality for all participants was protected. Parental consent was attained for all participants after parents were provided with full knowledge of all interventions taking place during this study and the summer program. Deception was not used.
RESULTS

Total Written Words, Words Spelled Correctly, and Correct Word Sequences

The students in the Three and Seven Point Cohorts showed improved in slope from baseline to treatment on TWW, WSC, and CWS while the Five Point Cohort did not. However, all of the cohorts increased on these indices when the average scores were manually computed. The Three Point Cohort improved from baseline to treatment as follows: TWW from $x = 20$ to 24, WSC from $x = 16$ to 18, and CWS from $x = 12$ to 12. The Five Point Cohort increased on TWW from $x = 30$ to 38, WSC from $x = 22$ to 27, and CWS from $x = 16$ to 19. Finally, the Seven Point Cohort improved from $x = 22$ to 44 on TWW, from $x = 17$ to 33 on WSC, and from $x = 14$ to 22 on CWS. As the intervention was designed to increase the number of words the students produced, these increases were expected. The students created a more detailed plan for their stories with the story map and with the assistance of their teachers. They also needed to write 30 percent more words during treatment than they did during baseline to earn reinforcement. Consequently, improvements in the number of words written also lead to increases in the number of correctly spelled words and the number of word sequences that the students wrote correctly.

In contrast, the slope of the Five Point Cohort showed some growth during treatment and then decreased to near or below baseline levels. Upon examination of the students’ individual performances, two students improved during treatment and two remained stable across baseline and treatment. Additionally, another student wrote an average of 69 words during baseline and was unable to write 30 percent more words in order to receive reinforcement. Finally, the last student strongly disliked writing and became more discouraged throughout the intervention, which lead to the student crying and/or refusing to complete the task.
Figure 1: Total Written Words

Three Point Cohort

Five Point Cohort

Seven Point Cohort

Session
Figure 2: Words Spelled Correctly

Three Point Cohort

Five Point Cohort

Seven Point Cohort
Figure 3: Correct Word Sequences

Three Point Cohort

Five Point Cohort

Seven Point Cohort
Correct Punctuation Marks

While the intervention led to improvements in the amount of words that the students wrote, and thus, increases in WSC and CWS, the intervention did not specifically target punctuation nor did it provide the students with feedback on their correct and incorrect usage of punctuation marks. Thus, as students increased their production, they did not show similar increases in CPM. The average CPM for the Three Point Cohort during baseline was 1 and it remained 1 during intervention. The average CPM was 3 during baseline and 2 during treatment for the Five Point Cohort. The Seven Point Cohort averaged 2 CPM during both phases.

As such, the slope of the students’ performance on this measure remained fairly constant throughout baseline and treatment in the Three Point Cohort. Assessment of the slope of the students’ individual performances showed that four of the six students in the Five Point Cohort used between zero and three CPM consistently throughout both baseline and treatment. The other two students applied more punctuation marks throughout the intervention, but did not use them consistently. Additionally, these students did not complete as many days of treatment as the other four students and this is reflected in the drop in the mean score on the last day of treatment. The variability in the slope for the Seven Point Cohort is due to the performance of two children. While two students used between zero and four CPM throughout baseline and treatment, these two students used between five and ten CPM during several sessions.
Figure 4: Correct Punctuation Marks

Three Point Cohort

Five Point Cohort

Seven Point Cohort
Percentage of Words Spelled Correctly and Percentage of Correct Word Sequences

The slope of the students’ performance on %WSC and %CWS remained relatively stable across baseline and treatment. On %WSC, the Three Point Cohort scored an average of 79 during baseline and 77 during intervention, the Five Point Cohort scored an average of 67 and 66, and the Seven Point Cohort averaged 72 and 73. On %CWS, the following scores were attained during baseline and intervention: Three Point Cohort $x = 58$ and $x = 51$, Five Point Cohort $x = 43$ and $x = 38$, and Seven Point Cohort $x = 57$ and $x = 48$. As the intervention did not focus on increasing the students accuracy, but rather focused on increasing their production, this level of performance was expected on these indicies.
Figure 5: Percentage of Words Spelled Correctly

Three Point Cohort

Five Point Cohort

Seven Point Cohort
Figure 6: Percentage of Correct Word Sequences

Three Point Cohort

Five Point Cohort

Seven Point Cohort

Session
Figure 7: Correct Minus Incorrect Word Sequences

Three Point Cohort

Baseline

Treatment

Five Point Cohort

Total of Correct Minus Incorrect Word Sequences

Seven Point Cohort

Session
Correct Minus Incorrect Word Sequences

The students in this study struggled to use correct punctuation consistently throughout baseline and treatment. Additionally, they were targeted for the summer program due to their difficulties in reading, which was often attributed to their lack of phonemic awareness. As such, these students had significant difficulties with their spelling skills. Since spelling and major punctuation greatly influence the number of correct and incorrect word sequences, the students often received negative scores on the CMIWS index. The Three Point Cohort scored an average of 5 during baseline and -1 during treatment, the Five Point Cohort scored an average of 4 and -2, and the Seven Point Cohort scored an average of 6 and 1. The variability of their performance may be attributable to the fact that the students typically used the same word several times incorrectly in a writing passage and that they were more likely to use challenging words based on the writing prompt they were given.

Inter-Rater Agreement

Inter-rater agreement was attained for 33 percent of the writing probes. In 2006, inter-rater agreement was assessed for TWW at 99.8 percent, CWS at 88.9 percent and IWS at 85.0 percent. In 2008, agreement was assessed for TWW at 96.1 percent, WSC at 98.1 percent, CPM at 95.4 percent, CWS at 90.6 percent, and IWS at 95.0 percent. When the scores were collapsed across both 2006 and 2008, agreement was assessed for TWW at 98.1 percent, CWS at 89.7 percent and IWS at 89.4 percent.
DISCUSSION

The multiple baseline design has many advantages in applied settings (Regan et al., 2005; Albertson & Billingsley, 2001; Lienemann et al., 2006). For instance, it shows differences in level, variability, and trends. It also allows determinations to be made based on the practical value of treatments by allowing teachers and parents to use visual inspection to monitor treatment effects. Furthermore, Smith, Best, Stubbs, Archibald, and Roberson-Nay’s (2002) comparison of “hard” and “soft” science journals showed that visual inscriptions, such as line graphs or charts, was the most important characteristic of hard sciences. Additionally, multiple baseline designs allow for intrasubject comparisons, which make comparisons between a child’s performance before, during, and after treatment possible (Morgan & Morgan, 2001). Lastly, Horner, Carr, McGee, Odom, and Wolery (2005) showed that this design is valid for establishing internal validity.

This study had the students draw a picture about their story for five minutes during baseline. This was done to ensure that there was an equal amount of planning time during both baseline and treatment as it was believed that the students were unlikely to think for a full five minutes about the story that they were going to write. However, the students were able to draw a picture for the whole five minute period. Thus, the students planned using the drawing for the same amount of time as they planned using the story map. This helped to ensure that the differences evident between the baseline and intervention phases were not due to differences in planning time, but due to differences in the type of planning that the students did.

Limitations

The primary limitation in this study occurred in order to remedy a confound in the 2006 methodology. Thus, there was a slight difference in the implementation of the treatment phase during the 2006 and 2008 summer programs. In 2006, the students were instructed that they would receive reinforcement based upon writing more words than they had ever written in the past sessions. However, several of the students maxed out and were unable to beat their previous scores. In order to ensure that all students were able to receive reinforcement by reaching their goal for TWW, the goal in 2008 was determined by finding the baseline mean and increasing it by 30 percent.

An additional limitation resulted from the overarching goal of the summer program. The students were referred for intervention based upon their reading deficits. Thus, a large portion of
the summer program focused on remediating these deficits, and these interventions ran concurrently with the writing intervention. It is possible that the some of the improvements seen in the students’ written expression skills were due to their improvements in reading and understanding of phonemic principles. However, the students were selected into their multiple baseline cohorts based upon their having a stable or declining baseline. Thus, although there may have been a cumulative effect of receiving reading and writing intervention, it is believed that the treatment effects are mostly due to the writing intervention.

There were also some differences in scoring of the CBM indices across raters. In order to correct for this, the raters received additional instruction in scoring, common errors were posted in the classroom, and the raters were allowed to score their students’ passage with the help of the primary investigator. Discrepancies in the scoring greatly decreased after the first three days of the study.

A final limitation is that the students may not be able to generalize the skills they gained during treatment using the story map if they do not have access to the graphic organizer. While the students may remember what they learned during the intervention, treatment affects were not monitored after treatment was discontinued due to time constraints. Thus, it is impossible to determine if the students maintained the improvements that they made or if these improvements were generalized to different writing tasks both with and without a graphic organizer.

Future Research

In general, the results for the Three Point Cohort were better and less variable than those of the Five and Seven Point Cohorts. This may be an artifact of data collection as the students with the most stable baseline were chosen to begin treatment first. It may also show that there were some practice effects for the students in the Five and Seven Point Cohorts. As these differences were very noticeable across several CBM indices, further research should seek to identify the implications that this might have on data collection in a Response to Intervention service delivery model. It should also seek to determine how much baseline is needed and whether there are significant differences in students who need more time in order to attain a stable baseline trend.

While the results showed that many of the students improved in production dependent indices: TWW, WSC, and CWS, they did not show these same improvements in CPM, which is another production dependent measure or in the production independent and combined indices.
However, the treatment was designed to assess general treatment affects and not to improve accuracy. Thus, it was very successful at this task and the stability of the students’ performance on %WSC and %CWS across baseline and treatment was expected. This shows that improving the quantity of output does not necessarily affect other aspects of writing, and that these skills may need to be directly taught. Future studies may include instruction in written expression skills including: spelling, grammar rules and/or punctuation in order to determine how this affects students’ performance on these measures.
REFERENCES


Appendix A: Story Starters

- John and Sara were lost in the jungle.
- One day Mary brought her pet skunk to school.
- Jill got a surprise package in the mail.
- One day I met a talking dog.
- My family is taking a trip to King’s Island.
- My friend and I landed on a deserted island.
- Jim and Fran spent the night in a cave.
- I found a secret door!
- I met an alien yesterday.
- This morning I woke up in a field.
- I have a magic wand in my pocket.
- I had the most unusual day.
- My friend found a time machine.
- I just saw a monster.
- I met the meanest pirate today.
Appendix B: Baseline Script

| Baseline: Distribute journal. Tell the student to write the date at the top of page. Write story starter on top line. Then say, |
| \[ \text{"I want you to write a story. Here is the first sentence... (Read assigned story starter)." Think about what would happen if (repeat story starter), then write as many sentences as you can in five minutes. If you don't know how to spell a word, just do the best you can. Write as many sentences as you can. Do your best writing.}" ] |
| \[ \text{"In order to write your story, I am going to ask you to draw a picture of your story so that I can tell who is in the story, when it takes place, and what happens."} \] |
| Use a stopwatch to time for 5 minutes of drawing time. |
| \[ \text{"Now, use this drawing to write your story. Write in complete sentences and do your best writing."} \] |
| Use a stopwatch to time for 5 minutes of writing time. Do not provide assistance during writing. If student asks for help, respond with "Just do your best work." |
| \[ \text{"Time. Stop and put your pencil down."} \] |
| Complete Drawing/Story Mapping Rubric by checking off which parts were included in the student's drawing. |
| Score the story for TWW, WSC, CPM, CWS, and IWS. |
Appendix C: Treatment Script

**Story Mapping + Incentive:**

Distribute journal. Tell the student to write the date at the top of page. Write story starter on the top line, and the goal (30% above TWW baseline mean) on the top left corner. Then say,

"I want you to write a story. Here is the first sentence... (Read assigned story starter). Think about what would happen if (repeat story starter), then write as many sentences as you can in five minutes. If you don’t know how to spell a word, just do the best you can. Write as many sentences as you can. Do your best writing.

"I want you to write ____ words (point to top left of student’s page, where the goal is written). If you can beat that score, you will earn a special SNACK COUPON. You can exchange these snack coupons for very special treats to eat.

“In order to meet your goal, I am also going to help you create a story map. A story map tells you some of the story, such as who is in the story, when it takes place, and what happens.”

Use a stopwatch to time for 5 minutes of story mapping time. Ask student the following questions and ensure that the student writes something in each box on the story map.

Who is your story about?
When/where did your story happen?
What happened in your story?
What was the problem?
How did (characters name) feel?
How did the story end?

Questions may vary depending on the student's response.

“Now, use this map to write your story. Write in complete sentences and do your best writing. If you beat your best score, you will earn a snack coupon.”

Use a stopwatch to time for 5 minutes of writing time. Do not provide assistance during writing. If student asks for help, respond with “Just do your best work.”

“Time. Stop and put your pencil down.”

Complete Drawing/Story Mapping Rubric by checking off which parts were included in the student's story map.

Score the story for TWW, WSC, CPM, CWS, and IWS.
Appendix D: Story Map

**Characters**

____________________
____________________
____________________
____________________

**Setting**

Time:________________

Place:________________

**The Problem**

__________________________________________________
__________________________________________________
__________________________________________________
__________________________________________________

**The Goal**

___________________________________________________
___________________________________________________
___________________________________________________

**Action**

___________________________________________________
___________________________________________________
___________________________________________________
___________________________________________________

**The Outcome**

___________________________________________________
___________________________________________________
___________________________________________________
___________________________________________________
Appendix E: Scoring Rubric

Date: ___________  Session: ___________  Phase: ___________

Drawing/Story Mapping Rubric (Pre-writing)

Give one point for each occurrence of the following story components. Total possible should be based on a minimum of one occurrence per story component:

_____ Who (characters)  _____ Where (setting)
_____ When (setting)  _____ What (events, problems, outcomes)
_____ How (action)  _____ Why

Percentage Integrity: _____

TWW: ____  WSC: ____  CPM: ____  CWS: _______  IWS: ____
Appendix F: Scoring Handout

**Total Written Words (TWW):** TWW is the total number of written words regardless of spelling, capitalization and grammar. Units of letters separated by a space are considered a word. Numbers and symbols that are not spelled out are not included. Titles and words copied from the story starters are included.

**Words Spelled Correctly (WSC):** WSC is the total number of words that are spelled correctly within the passage context. For example, “I have went their before.” would count as 4 WSC because although “their” is spelled correctly, it is incorrect within the context of the sentence. Words combined incorrectly (secondgrade) are counted as 1 word incorrect. Words that are incorrectly separated (tele vision) are counted as 2 incorrect words.

**Correct Punctuation Marks (CPM):** CPM is the total number of correctly applied punctuation marks.

**Correct Word Sequences (CWS):** CWS is the total number of sequences or two adjacent, correctly spelled words that are acceptable within the context of the phrase by someone whose first language is English. Capitalization and major punctuation (.!?) is included.

For example, the sentence “Her go to the stor at 4 o’clock yesterday.” would be scored:

\[ \wedge \text{Her} \wedge \text{go} \wedge \text{to} \wedge \text{the} \wedge \text{stor} \wedge \text{at} \wedge 4 \wedge \text{o’clock} \wedge \text{yesterday} \wedge . \]

The CWS for this sentence is four. “Her” shows a capitalization at the beginning of the sentence and is therefore, counted as a CWS. “Her go” is not syntactically correct and is scored as an incorrect word sequence (IWS). However, “go to” and “to the” are acceptable two word sequences. Since “stor” is spelled incorrectly, both “the stor” and “stor at” are IWS. “at 4,” “4 o’clock,” and “o’clock yesterday” are CWS. “yesterday.” is counted as an additional CWS because the period is a correctly used major punctuation mark.

**Incorrect Word Sequences (IWS):** IWS is the total number of sequences that do not fit the criteria listed above for CWS.

* The first written word should be capitalized.
* Proper names do not have to be capitalized to be spelled correctly. However, word sequences should be scored as follows:

  \[ \wedge \text{I will go to} \wedge \text{ohio} \wedge . \]