Learning Disabilities: A Contemporary Journal 11(2), 51-64, 2013

Analyzing the Effects of Story Mapping on the Reading Comprehension of Children with Low Intellectual Abilities

Matthias Grünke*
University of Cologne, Germany

Jürgen Wilbert
University of Potsdam, Germany

Kim Calder Stegemann
Thompson Rivers University, USA

This single-case study examined the effects of a graphic organizing strategy on the ability of children to improve their text comprehension abilities. Participants were six students between ten and fourteen years old with major problems in understanding what they read. The intervention intended to teach them to visually highlight key elements of a passage, and thus, to deepen their understanding of it (story mapping). An AB multiple baseline design across subjects was applied. The intervention points were randomly determined within a preset range for each participant. In accordance with the emerging trend to apply inferential statistics as a supplement to visual inspection and the calculation of effect size measures, a randomization test and a piecewise regression procedure were used to analyze the data. Results suggested that the story mapping technique was very beneficial in improving reading comprehension of struggling learners. The potentials of the intervention as well as of the statistical tests in analyzing data from single-case studies are discussed.

Keywords: **Text comprehension, story mapping, single-case analysis, randomization test, piecewise-regression analysis**

Introduction

Reading comprehension is the ability to construct and extract meaning from a written text (Mastropieri & Scruggs, 1997). It is considered to be the most critical skill that is needed to succeed in school. If readers have serious difficulties to gather relevant information from a historical account, a mathematical word problem, or a passage in a biology book, they are bound to fail in most every task that is put before them. To be able to understand a written text, students must be proficient in lower levels of reading (phonemic awareness, phonics, fluency, and vocabulary) (National Reading Panel, 2003; RAND Reading Study Group, 2002). In addition, they have to hold in their working memory a mental model of a circumstance, event, or problem being described. Readers need to revise existing understanding of a certain matter while gathering new information (Blanc, Kendeou, Van den Broek, & Brouillett, 2008). They must make connections to their prior knowledge. In order to do so,

*Please send correspondence to: Matthias Grünke, matthias.gruenke@uni-koeln.de.

it is helpful to have a broad knowledge base to fall back on (Van den Broek, Rapp, & Kendeau, 2005). Sometimes, ideas in a text are unfamiliar or not well defined. In such cases, students have to be able to bridge conceptual gaps. Finally, they need to be familiar with different text structures and must know how ideas are organized in either narrative or expository material (McCormick & Zutell, 2011).

Most children develop sufficient comprehension abilities until they reach 3rd grade. This happens even though most teachers just focus on basic skills like phonological awareness, decoding, and fluency (Whitehurst & Lonigan, 1998). Proficient readers use effective comprehension strategies without being taught, and without being aware of implementing them (Swanson & De La Paz, 1998). However, a considerable number of students do not acquire the necessary skills to derive meaning from written material, even though they do not exhibit problems in decoding (Lipka, Lesaux, & Siegel, 2006). This might be due to insufficient abilities to make inferences, draw conclusions, recall and summarize information, actively monitor their comprehension, to a limited working memory capacity, to a lack of prior knowledge, as well as to other reasons (Adlof, Catts, & Lee, 2010; Catts, Adlof, & Weismer, 2006).

Fortunately, there are many evidence-based practices to improve reading comprehension in students, who are struggling. Reed and Vaughn (2012) list the following approaches, which have all proven to be helpful: (1) Teaching relevant background knowledge (like definitions of unknown vocabulary, translation of foreign phrases, clarification of difficult concepts, etc.), (2) outlining different kinds of text structure, (3) helping to identify the main ideas in a text, (4) demonstrating how to summarize a text by making inferences and synthesizing the information, and (5) using an instructional activity called reciprocal teaching, where a student and a teacher (or a tutor) engage in a dialogue concerning different parts of a text in order to construct the meaning. Another effective and well-known approach is the use of graphic organizers. These are visual learning strategies that make the structure of concepts as well as relationships between them apparent. They help students to create an organized schema and to connect prior knowledge with the content of a text that a learner is reading (Shanahan, Callison, Carriere, Duke, Pearson, Schatschneider, & Torgesen, 2010). By using these tools, a child can reduce the amount of semantic information he or she needs to process in order to extract meaning (Faggella-Luby, Schumaker, & Deshler, 2007; Jitendra & Gajria, 2011). Graphic organizers thus reduce working memory overload (O'Donnell, Dansereau, & Hall, 2002).

Among all the different kinds of graphic organizers (semantic maps, concept maps, semantic feature analysis, Venn diagrams, etc.), story maps are probably the ones used most widely. With these tools, the teacher can model for the students how to locate the elements (settings, characters, problems, events, solutions, and conclusions) of a narrative. He or she writes the relevant information into a visual depiction, while thinking aloud. Graph number 1 shows a typical story map.

But even though story maps and the other approaches mentioned above have proven to be effective as a whole (e. g. Kim, Linan-Thompson, & Misquitta, 2012; Sencibaugh, 2007), looking at the findings in detail offers a rather ambivalent picture. Apparently, using the same intervention with different students showing comprehension problems does not result in improvements of similar magnitude. Watson, Gable, Gear, and Hughes (2012) reason that what is beneficial for one particular reader is not necessarily very profitable for another. One student might have problems with vocabulary, another with making inferences, and a third one with finding the main idea. They all need different kinds of interventions, focusing on different goals. In addition, strategies which are effective for younger learners may not be useful for older students. Further, it should be noted that even research-based approaches can yield poor results if applied in an inconsistent or highly modified manner (Kim, Linan-Thompson, & Misquitta, 2012).

In this paper, we evaluated the effectiveness of a story mapping procedure with a small group of subjects who seemed to be especially receptive to benefit from this approach. Previous research suggests that graphic organizers like story maps are particularly helpful for prepubescent students with rather low general intellectual abilities and low comprehension skills, but with a sound proficiency in reading fluently (Grünke, 2011). Children with these characteristics oftentimes struggle to find the main idea in a text or to grasp its overall theme. This is a scenario where story mapping seems to be particularly useful. We thus selected a group that met the aforementioned description. Using a single-case design, we investigated whether applying the story mapping technique with these kind of children yields especially great treatment effects.

Method

Subjects and Setting

The study was conducted in Germany. Three 5th grade students from a regular education public school and three 8th grade students from a school for children with learning difficulties served as subjects. Four of them were female (Anna, Bella, Christina, and Dunja), two of them were male (Egor and Fabian) (names altered, for anonymity). The girls were 11, 10, 14, and 14 years old, the boys were 11 and 13 years old. According to their teacher, the three students from the school for children with learning difficulties (13, 14, and 14 years old) were approximately three years behind in their emotional development and behaved generally very childlike. Bella's parents were from Kazakhstan, Dunja's parents from Serbia, and Egor's parents from Russia (the remaining children did not have an immigrant background). The schools that the subjects attended were located in a major city in North Rhine-Westphalia, Germany. All students were identified by their teachers as having outstanding difficulties in text comprehension despite an adequate ability to read fluently. A screening using the German Reading Comprehension Test for First to Six Graders (ELFE 1-6, Lenhard & Schneider, 2006) revealed that all six children possessed reading comprehension skills below the 25th percentile of 4th graders. The general intellectual abilities of the subjects as measured with the German Number Combination Test (ZVT, Oswald & Roth, 1987) ranged also in the lowest quartile.

Observation and intervention occurred in separate rooms of the two schools during a daily period of independent class work, which was still in progress when the subjects returned.

Dependent Variable and Measurement Procedures

We selected 18 narratives from three different German story books (Wölfel, 1974; 2010a; 2010b). All of them were altered in a way that it was possible to formulate exactly ten comprehension questions about each tale that covered its main content. Sometimes, additional information had to be added, sometimes, information had to be eliminated to make narratives comparable. The comprehension questions were stated in a way that only one specific and distinct answer was possible to be counted as correct. Subsequently, we standardized the texts, so that each of them consisted of exactly 150 words. In a preliminary survey, the stories and comprehension questions were presented to twenty low achieving children between 9 and 10 years old in order to identify items that were either too easy or too hard to solve. We involved the insights from this preliminary survey to compose the final version of our question sets.

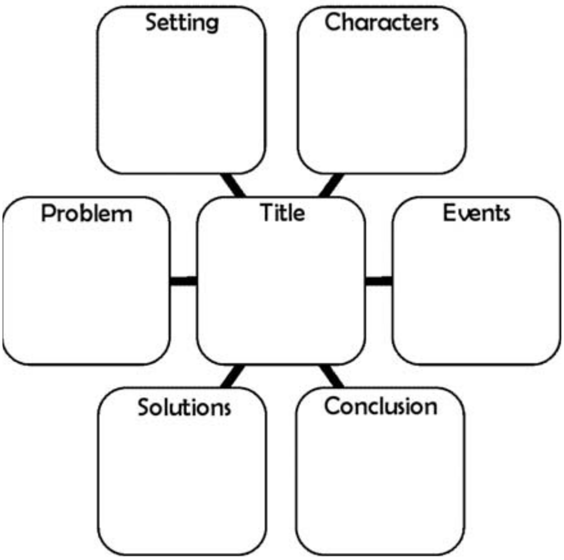
In the course of the study, each student was individually presented with a different story and a different set of comprehension questions for 18 consecutive school days. The order of the tales was randomly chosen for each child. Each student was asked to read a respective story out loud and then to write down the answers to the corresponding questions on a worksheet. While reading, the children were allowed to do whatever seemed meaningful to them to memorize the main content of each text (take notes, rehearse the information verbally, draw pictures, …). When the students decided to take the questionnaire, the sheet with the story as well as any aids (notes, pictures, story maps, …) were taken away. The children were given a time limit of 15 minutes to finish their daily assignment (reading a text, rehearsing its content, and answering the comprehension questions).

Intervention

The teaching of the use of story mapping strategy was conducted by a male graduate student. Before working with the children, he was extensively prepared by the first author on how to instruct boys and girls to effectively apply this graphic organizing technique. Daily individual training for each child lasted 30 minutes. The student instructor used a German version of figure 1 for the intervention. Reading passages were taken from the aforementioned story books (Wölfel, 1974; 2010a; 2010b). Of course, the 18 narratives that were used to measure the children's performance were exempt.

To teach the boys and girls how to better comprehend narrative texts by using a story map, the student instructor followed a procedure outlined by Idol (1987): (1) Modeling phase (the teacher demonstrates how a story map is used by reading a tale out loud and by stopping whenever important information is mentioned to fill out parts of his or her worksheet), (2) lead phase (the children read stories independently and complete their maps while the teacher prompts and encourages them to review their work results and to add details that they might have overlooked), (3) test phase (the children read texts, draw maps of their own, ask questions pertaining to the content, answer them, and fill in the components into their maps without close supervision by the teacher - the teacher only intervenes if the students ask for or obviously needs help). The first session always focused on the modeling phase. During the following two to three sessions, the lead phase took up the greatest share of instructional time. Depending on the skill level that a child had reached until then, the remaining sessions were either devoted entirely to the test phase or to other phases that still needed rehearsal.

Figure 1. **A Story Map**



To ensure that the intervention procedures were carried out as designed, the student instructor and the first author stayed continuously in close contact via e-mail and met on a weekly basis to evaluate past lessons and to discuss any further procedures.

Experimental Design

An AB multiple baseline design across subjects was applied. Usually, researchers continue with baseline observations until the baseline stabilizes. But this procedure constitutes a threat to the internal validity of a study. It creates a bias in favor of identifying an intervention effect where none exists. There is no way of knowing what the baseline would have looked like if it had continued for a little longer. A couple of upward random variations followed by a couple of downward random variations could easily and wrongly be interpreted as stabilization of the baseline (Todman, 2002). An alternative to wait until a baseline stabilizes would be to come up with a preset number of total probes and a minimum number of baseline sessions as well as a minimum number of intervention sessions, and then to determine the beginning of the treatment by chance. This procedure cannot avoid random variations, but it turns potentially systematic nuisance variables into random nuisance variables, and thus increases the internal validity of the findings. In the present study, a total number of 18 daily observations was determined for the baseline and the intervention sessions. It was previously decided that the baseline phase had to consist of at least three, but not more than eight probes. This yielded six possible intervention points for each subject (the treatment could either start after the third, the fourth, the fifth, the sixth, the seventh, or the eighth baseline observation). According to a random drawing of these options, teaching the story mapping technique started for Anna after the fourth, for Bella after the seventh, for Christina after the fourth, for Dunja after the sixth, for Egor after the fifth, and for Fabian after the eighth baseline probe.

Results

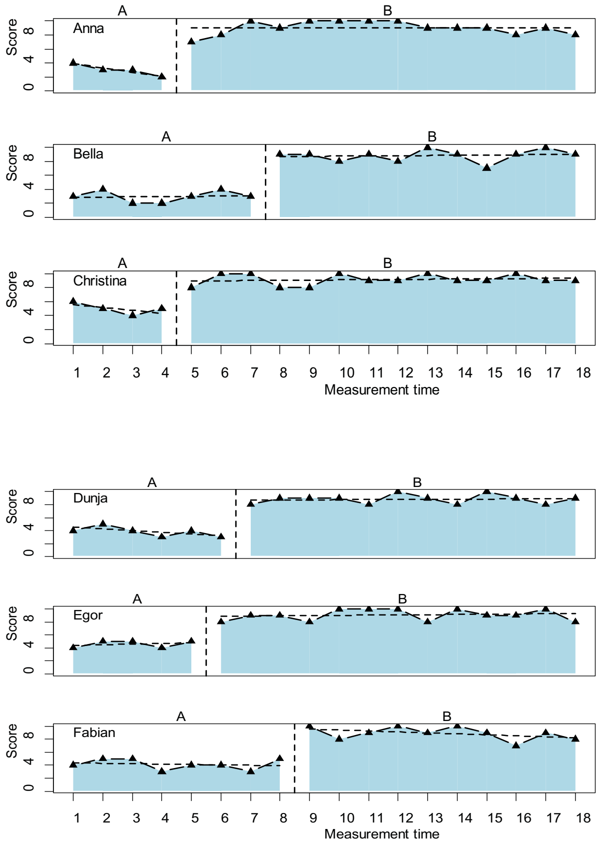
Information on the number of correctly answered comprehension questions is reported in Table 1.

Table 1. ***Correctly answered comprehension questions***

Student		Baseline	Intervention
Anna	N (Probes)	4	14
	Raw Scores	4; 3; 3; 2;	7; 8; 10; 9; 10; 10; 10; 10; 9; 9; 9; 8; 9; 8;
	M	3.00	9.00
Bella	N (Probes)	7	11
	Raw Scores	3; 4; 2; 2; 3; 4; 3;	9; 9; 8; 9; 8; 10; 9; 7; 9; 10; 9;
	M	3.00	8.82
Christina	N (Probes)	4	14
	Raw Scores	6; 5; 4; 5;	8; 10; 10; 8; 8; 10; 9; 9; 10; 9; 9; 10; 9; 9;
	M	5.00	9.14
Dunja	N (Probes)	6	12
	Raw Scores	4; 5; 4; 3; 4; 3;	8; 9; 9; 9; 8; 10; 9; 8; 10; 9; 8; 9;
	M	3.83	8.83
Egor	N (Probes)	5	13
	Raw Scores	4; 5; 5; 4; 5;	8; 9; 9; 8; 10; 10; 10; 8; 10; 9; 9; 10; 8;
	M	4.60	9.08
Fabian	N (Probes)	8	10
	Raw Scores	4; 5; 5; 3; 4; 4; 3; 5;	10; 8; 9; 10; 9; 10; 9; 7; 9; 8;
	M	4.13	8.90

For a first rough estimation of the data, we conducted a visual inspection by considering slopes, phase changes, and variability in the measure set (Gast & Spriggs, 2010). Figure 2 maps the performance progress for all students including reference lines depicting the slope of both phases. The variable within students and phases is considerably small compared to differences between students and phases. We found a uniform increase of performance with the beginning of the B-phase for all students, while the slope lines did not show a consistent increase in the B-phase.

Figure 2. **Number of correctly answered questions for each student across the course of the study. The vertical dashed line indicates the beginning of the intervention phase. The horizontal dashed lines depict the slope parameter of a linear regression of the A and B-phase for each student**



In addition, non-overlapping indices were calculated for all participants (Parker, Vannest, & Davis, 2011) as a means to measure the strengths of the treatment outcomes (effect sizes). Table 2 shows no overlap of data for any applied measure. Percentage of non-overlapping data, non-overlap of all pairs, percentage exceeding the median, and percentage of all non-overlapping data were all 100%.

Table 2. ***Descriptive statistics for the six single-cases and two aggregations. The first aggregation results from an interpolation or summing up of the values of the six cases. The second aggregation is based on a procedure described in Wilbert (2014). The subscripted characters refer to the respective measurement phase.***

	Case							
statistics	Anna	Bella	Christina	Dunja	Egor	Fabian	Aggregation (weighted average/sum)	Aggregation (overlapping)
n_A	4	7	4	6	5	8	34	34
n_B	14	11	14	12	13	10	74	74
M_A	3.0	3.0	5.0	3.8	4.6	4.1	3.9	0.0
M_B	9.0	8.8	9.1	8.8	9.1	8.9	9.0	5.0
$M_B - M_A$	6.0	5.8	4.1	5.0	4.5	4.8	5.0	5.0
min_A	2	2	4	3	4	3	2	-1.1
min_B	7	7	8	8	8	7	7	2.9
max_A	4	4	6	5	5	5	6	1.2
max_B	10	10	10	10	10	10	10	7
SD_A	0.8	0.8	0.8	0.8	0.5	0.8	0.8	0.7
SD_B	1.0	0.9	0.8	0.7	0.9	1.0	0.9	1.1
ac_A	0.0	0.0	0.0	0.0	-0.5	-0.2	-0.1	0.2
ac_B	0.3	-0.2	-0.2	-0.4	-0.3	-0.1	-0.1	0.1
b_A	-0.6	0.0	-0.4	-0.3	0.1	-0.1	-0.2	-0.1
b_B	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0
b_{AB}	0.3	0.5	0.2	0.4	0.3	0.4	0.4	0.3
$b_B - b_A$	0.6	0.0	0.4	0.3	-0.1	-0.1	0.2	0.1
PND	100	100	100	100	100	100	100	100
PEM	100	100	100	100	100	100	100	100
NAP	100	100	100	100	100	100	100	100
PAND	100	100	100	100	100	100	100	100

Note. *n* = number of data points, *M* = mean, *SD* = standard deviation, *ac* = lag one autocorrelation, *b* = slope of a linear regression, *PND* = percent non-overlapping data, *PEM* = percent exceeding the median, *NAP* = non-overlap of all pairs, *PAND* = percent all non-overlapping data.

Supplementally, we analyzed the data using inferential statistics. It is becoming a common standard in single-subject research to not rely only on visual inspection and effect size measures when drawing inferences from case studies. Unfortunately, most of the usual parametric tests are unsuitable for this purpose. One of the major objections in this respect are statistical problems caused by auto-correlated data. When dealing with AB designs, however, randomization tests (e. g. Dugard, File, & Todman, 2011) and piecewise regression analysis (e. g. Center, Skiba, & Casey, 1986) have proven to be very helpful approaches in a lot of instances where data from case studies had to be statistically analyzed. Explaining how these procedures can be applied in single-subject research would go beyond the scope of this paper. We thus refer the reader to the above mentioned literacy sources for greater details.

In order to decide if at least one of these two strategies is suitable for our purposes, we computed an empirical power and alpha-error estimation of the randomization test and of the piecewise-regression analysis for the given structure and distribution of our data. The empirical power and alpha-error analyses were based on a Monte Carlo study. We simulated 2000 data sets with the same distributions of parameters that were prevalent in the observed data. Thereby we assumed that all the effects that we found in the data were systematic and not random. We subsequently computed statistical tests on the simulated data to estimate their power. In a second step, we simulated new data sets with specific effects set to zero. We thereby produced data sets with the same structure, but without level or slope effects. Furthermore, we conducted statistical analyses on these data sets to estimate the proportion of false positive results that the tests produce under the given circumstances for level and slope effects. The proportion of false positive results is an estimantion for the alpha-error probability of the method of analysis that we used.

All calculations were carried out with an R-package (SCDA, Wilbert, 2014) which contains a convenient function for this procedure. The random data generating model assumed the following parameters: a six cases multiple baseline with a B-phase beginning at the 5^{th}, 7^{th}, 9^{th}, 5^{th}, 8^{th}, and 6^{th} measurement-point and each case with a total of 18 measurements. The underlying distribution of the measured values was set to $M = 3.90$ ($SD = 0.80$) and we assumed a reliability of measurement of $r_{tt} = .80$. The effects of the intervention were estimated $d_{level} = 6.52$, $d_{slope} = 0.10$, and $d_{trend} = -0.10$. Table 3 depicts the resulting power and alpha-error. Both randomization and regression analysis have a very high power (both 100%) and low alpha-error (9.2% and 5.6%, respectively), when estimating a level effect due to the intervention. However, the slope-effect was far too small and the measurements were too little for a sufficient power of the analysis (6.6% for the randomization test and 45.4% for the regression model). These results suggest that a statistical test on significance of the slope effect should not to be perfomed.

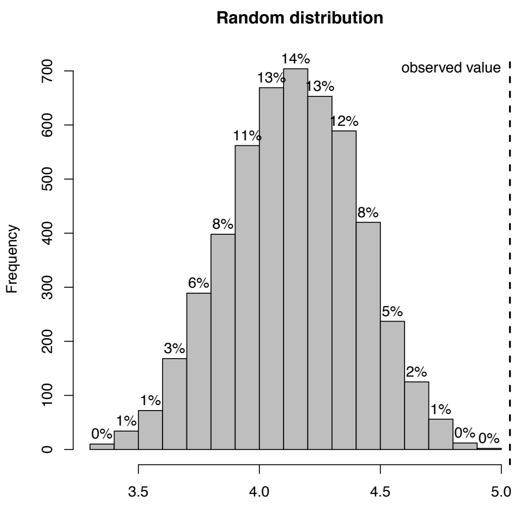
Table 3. ***Power and alpha-error of randomization test and piecewise-regression model analyses for level and slope effects based on the parameters of the study at hand.***

	Power	Alpha-error
Randomization Test Level	100.0	9.2
Randomization Test Slope	6.6	2.8
PLM Level	100.0	5.6
PLM Slope	45.4	3.4

Based on the results of the power analysis we decided to conduct both a randomization test and a piecewise-regression analysis. For the randomization test, we computed the difference of the mean values of the two phases (M_B - M_A) as the target statistic under randomly varying combinations of starting positions of the B-Phase (see above). The actual observed mean difference was then compared to the resulting distribution (see Figure 3). The percentile of the observed value within this distribution is the resulting p-value of the test.

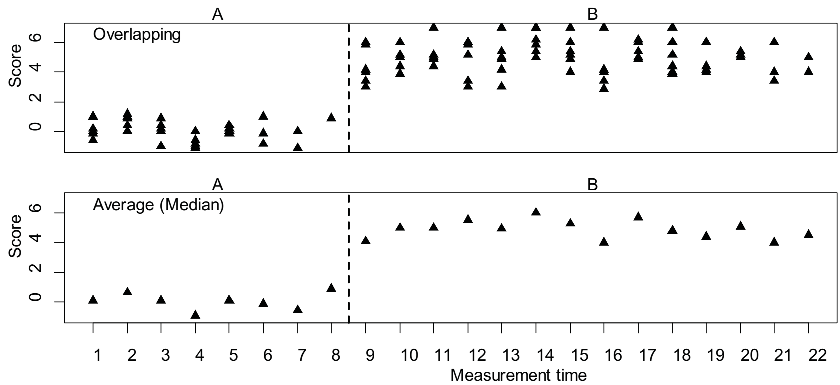
The random sample was created based on the assumption that the intervention could have started after the third, the fourth, the fifth, the sixth, the seventh, or eighth baseline observation. From the resulting possible 46,656 combinations of starting points of the B-phase, a sample of 5,000 was drawn. All mean differences (MB – MA) for this 5,000 random combinations were below the observed value of MB - MA = 5.04 (distribution of mean differences: M = 4.13, SD = 0.27, min = 3.31, max = 3.37) giving a $p < .0002$ (assuming a normal distribution of the mean differences: $z = 3.31$, $p < 0.001$).

***Figure 3*. Reference distribution of the randomization test**



In order to carry out a piecewise-regression analysis, we firstly aggregated the six single cases to one single case following the procedure described in Wilbert (2014) and using the SCDA-package. This was done by centering the data of all cases on the mean of the A-phase of each single-case. In a second step, we sorted the values of all A-phases by their measuring time and merged them into one single A-phase. We then did the same with the values of the B-phases. Finally, the measuring times of the B-phases were increased to start one measurement after the last measurement of the A-phase. We subsequently recombined the resulting merged A- and B-phase into a new single-case including the measurements of all the cases. This aggregation allowed for a combined analysis of all six single-cases. Figure 4 depicts the resulting aggregated case.

Figure 4. **Plot of the aggregated data of the six single-cases. In the lower figure, multiple measurements for one measurement point are median-averaged**



The complete regression model had a significant fit ($F[3, 104] = 199.12$; $p < 0.001$; adjusted $R^2 = .85$). Applying an auto-regression model based on a general least square estimation led to a negligible difference in model parameters (see Table 4).

Table 4. ***Piecewise-regression model for the aggregated six single cases. ΔR^2 is computed with likelihood-ration test comparing the full model with the model without the target predictor (see Beretvas & Chung, 2008).***

	B	*SE*	*t*	*p*	ΔR^2
Intercept	0.28	0.36	0.79	.43	
Trend	-0.08	0.09	-0.90	.37	.001
Level	5.36	0.49	10.87	.00	.169
Slope	0.08	0.09	0.88	.38	.001

Discussion

Main Findings

The current study investigated the effects of a graphic organizing technique (story mapping) on the reading comprehension of six students between ten and fourteen years old, who had sufficient decoding abilities, but possessed rather limited intellectual skills and struggled with constructing and extracting meaning from a text. Results suggested that the strategy was extremely effective. All subjects were able to dramatically increase the number of correct responses in the continuously administered probes (from M = 3.88 during baseline to M = 8.97 during intervention). All applied procedures to measure the effectiveness of the treatment (visual inspection, effect size calculation, randomization test, piecewise regression analysis) indicated that teaching to use story maps has a tremendous potential to help children like the ones involved in our experiment to better understand a text. Previous studies were able to demonstrate that this method can be a powerful intervention (e. g. Babyak, Koorland, & Mathes, 2000; Boon, Fore III, Ayres, & Spencer, 2005; Boulineau, Fore III, Hagan-Burke, & Burke, 2004; Gardill & Jitendra, 1999; Lapp, Fisher, & Johnson, 2010; Smith, Boon, Stagliano, & Grünke, 2011). However, the effects that are outlined in these earlier research papers have never reached the magnitude of ours. We were thus able to demonstrate that story mapping can be particularly effective when it is used with children who are especially eligible for this kind of intervention.

Limitations

One often-raised concern with single-case studies is their purported limited generality. Because these designs include only a very small number of subjects, they are often considered to possess a rather constricted external validity. However, generality can easily be demonstrated via direct replication. As indicated above, there are already quite a number of studies that document the benefits of story mapping with struggling learners. Generality could certainly be established if more reports emerge that support the assumption that this technique is especially helpful with students who fit the criteria that we used to select our sample. Another objection to the explanatory power of this study is the reference to the rather specific type of text that was used. Students only worked with short stories (narratives) that were taken from books written by a certain author. It has yet to be determined whether story mapping is equally effective for expository texts.The comprehension questions for each story were obviously equally difficult to answer. There were only marginal variations in the scores during the baseline and the treatment phases for each student. Performance changes were apparently due to whether or not the children had received some instructions on how to use the strategy. As mentioned above, the order in which the stories were presented to each individual subject was randomly chosen. The increases in correctly answered comprehension questions came about very abruptly. But even if they developed steadily over time, it would have not constituted a threat to the internal validity of the study.

One serious limitation of the study pertaining to the design was that no post treatment data were collected. Considering the large and instant treatment gains, it appears unlikely that the performance of the subjects would return anywhere close to base level upon finishing the intervention. However, this is just an assumption. No data is available to support this hypothesis.

In this study, we followed the trend of using inferential statistics as a supplement to the typical routine of analyzing data from case studies by just visually inspecting them or calculating effect sizes. One could argue that this undertaking was dispensable in our instance, because the effects were very apparent. As Edwards, Lindman, and Savage (1963) have commented on obvious treatment outcomes in single-subject studies, "… you know what the data mean when the conclusion hits you between the eyes" (p. 217). However, the conclusions that a researcher draws from a given data set are not always obvious. And even if they are for one person, this does not mean that someone else arrives at the same bottom line. Oftentimes, the interpretation of the findings seems to be left too much to the subjective discretion of the respective authors. Brossart, Parker, Olson, and Mahadevan (2006) point out that the inter-rater-reliabilities of visual inspections are remarkably low. Even if raters have been excessively trained in how to make sense of graphs depicting the course of the measurements, they still do not come up with very homogeneous interpretations. The line between an instance where a statistical analysis seems advisable and one where it seems redundant is virtually impossible to draw. Thus, it is reasonable to apply statistical tests when analyzing data from single-subject designs whenever possible.

Instances where this seems inappropriate are situations in which the robustness of such approaches has to be questioned due to high auto-correlations among original scores (e. g. Sierra, Solanes, & Quera, 2005). But as mentioned above, this did not constitute a threat to the internal validity of our study. The tests that we used are very immune to this jeopardy, when they are applied with data from an AB multiple baseline design across subjects (Wilbert, in press).

Practical Implications and Future Research

Teaching children to effectively extract meaning from a text is certainly one of the most important tasks that schools have to face. Without this ability, students will inevitably fail in their academic endeavors. In addition, they will miss out on a whole array of activities that make life enjoyable (like reading a book or communicating through social networking tools) and will struggle immensely in many of their daily routine activities (like understanding an instruction book, a letter from an agency, or the latest news on an iPad).

According to the findings of our study, helping children like the ones in our experiment to better extract meaning from a text through the use of story maps is anything but an insurmountable challenge. The student instructor who functioned as teacher to our sample did not receive extensive training prior to familiarizing the six boys and girls with the particular graphic organizing technique that we used. This experience raises hopes that this strategy could profitably be applied by a tutor on a one-to-one basis in a regular or inclusive classroom. Peer-tutoring has proven to be one of the most effective ways in a whole array of different academic content areas (Bowman-Parrot, Davis, Vannest, & Williams, 2013; Greenwood, Arreaga-Mayer, Utley, Gavin, & Terry, 2001). But finding effective procedures for struggling students that can easily carried out by their class mates, remains a great challenge. A lot of evidence-based interventions require a considerable amount of expertise on the side of the instructor. In contrast, story-mapping seems to be a very expedient tool to be used in peer-tutorial settings by children who do not necessarily have to possess an abundance of teaching skills. It can thus effectively contribute to break through the "… declining spiral of frustration, anxiety, and more failure" (Slavin, 2005, p. VIII) that students with low comprehension and low intellectual skills so often experience.

References

Adlof, S., Catts, H., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43*, 332-345.

Babyak, A. E., Koorland, M., & Mathes, P. G. (2000). The effects of story mapping instruction on the reading comprehension of students with behavioral disorders. *Behavioral Disorders, 25*, 239-258.

Beretvas, S., & Chung, H. (2008). An evaluation of modified R2-change effect size indices for single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 120-128.

Blanc, N., Kendeou, P., Van den Broek, P. W., & Brouillet, D. (2008). Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes, 45*, 103-121.

Boon, R. T., Fore III, C., Ayres, K., & Spencer, V. G. (2005). The effects of cognitive organizers to facilitate content-area learning for students with mild disabilities: A pilot study. *Journal of Instructional Psychology, 32*, 101-117.

Boulineau, T., Fore III, C., Hagan-Burke, S., & Burke, M. D. (2004). Use of story-mapping to increase the story-grammar text comrehension of elementary students with learning disabilities. *Learning Disability Quarterly, 27*, 105-121.

Bowman-Perrot, L., Davis, H., Vannest, K. J., Williams, L., Greenwood, C. R., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*, 39-59.

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*, 531-563.

Catts, H., Adlof, S., & Ellis W.S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293.

Center, B. A., Skiba, R. J., & Casey A. (1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Educational Science. 19*, 387-400.

Dugard, P., File, P., Todman, J. (2011). *Single-case and small-n experimental design*. New York, NY: Taylor & Francis.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193-242.

Faggella-Luby, M. N., Schumaker, J. S., & Deshler, D. D. (2007). Embedded learning strategy instruction: Story-structure pedagogy in heterogeneous secondary literature classes. *Learning Disability Quarterly, 30*, 131-147.

Gardill, M. C., & Jitendra, A. K. (1999). Advanced story map instruction: Effects on the reading comprehension of students with learning disabilities. *Journal of Special Education, 33*, 2-17.

Gast, D. L., & Spriggs, A. D. (2010). Visual analysis of graphic data. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 199-233). New York, NY: Routledge.

Greenwood, C. R., Arreaga-Mayer, Utley, C. A.,Gavin, K. M., & Terry, B. J. (2001). Class Wide PeerTutoring Learning Management System. *Remedial and Special Education, 22*, 34-47.

Grünke, M. (2011, March). *Fostering reading comprehension in children with learning problems through story-mapping*. Paper presented at the Spring Conference of the Association of Empirical Educational Research (AEPF) in Bamberg, Germany.

Idol, L. (1987). Group story mapping: Comprehension strategy for both skilled and unskilled readers. *Journal of Learning Disabilities, 20*, 196-205.

Jitendra, A., & Gajria, M. (2011). Reading comprehension instruction for students with learning disabilities. *Focus on Exceptional Children, 43*, 1-16.

Kim, W., Linan-Thompson, S, & Misquitta, R. (2012). Critical factors in reading comprehension instruction for students with learning disabilities: A research synthesis. *Learning Disabilities Research & Practice, 27*, 66-78.

Lapp, D., Fisher, D., & Johnson, K. (2010). Text mapping plus: Improving comprehension through supported retellings. *Journal of Adolescent & Adult Literacy, 53*, 423-426.

Lenhard, W., & Schneider, W. (2006). *Reading comprehension test for first to six graders (ELFE 1-6)*. Göttingen, Germany: Hogrefe.

Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade 4 students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*, 364-378.

Mastropieri, M. A., & Scruggs, T. E. (1997). Best practices in promoting reading comprehension in students with learning disabilities. *Remedial and Special Education, 18*, 197-213.

McCormick, S., & Zutell, J. (2011). *Instructing students who have literacy problems*. Boston, MA: Pearson.

National Reading Panel. (2003). *Put reading first: The research building blocks for teaching childrento read*. Washington, DC: National Institute of Child Health and Human Development.

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Education Psychology Review, 14*, 71-86.

Oswald, W.D., & Roth, W. (1987). *Number connection test (ZVT)*. Göttingen, Germany: Hogrefe.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*, 303-322

RAND Reading Study Group (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: Office of Education Research and Improvement.

Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading, 16*, 187-217.

Sencibaugh, J. M. (2007). Meta-analysis of reading comprehension interventions for students with learning disabilities: Strategies and implications. *Reading Improvement, 44*, 6-22.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Sierra, V., Solanas, A., & Quera, V. (2005). Randomization tests for systematic single-case designs are not always appropriate. *The Journal of Experimental Education, 73*, 140-160.

Slavin, R. (2005). Forword. In G.D. Sideridis, & T.A. Citro (Eds.), *Research to practice: Effective interventions in learning disabilities* (pp. VIII-IX). Weston, MA: LDW.

Smith, S., Boon, R. T., Stagliano, C., & Grünke, M. (2011). Story Mapping: Eine Methode zur Verbesserung der Fähigkeit von leseschwachen Grundschulkindern, Sachtexte zu verstehen. *Empirische Sonderpädagogik, 3*, 37-50.

Swanson, P. N., & De La Paz, S. (1998). Teaching effective comprehension strategies to students with learning and reading disabilities. *Intervention in School and Clinic, 33*, 209-218.

Todman, J. (2002). Randomisation in single-case experimental designs. *Advances in Clinical Neuroscience and Rehabilitation, 2*, 18-19.

Van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*, 299-316.

Watson, S. M. R., Gable, R. A. Gear, S. B., & Hughes, K. C. (2012). Evidence-based strategies for improving the reading comprehension of secondary students: Implications for students with learning disabilities. *Learning Disabilities Research & Practice, 27*, 79-89.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872.

Wilbert, J. (2014). *Using the SCDA package (0.8) for analyzing single and multiple case AB designs.* Retrieved April 5, 2014 from https://dl.dropboxusercontent.com/u/32192485/SCDA%20Documentation.pdf

Wilbert, J. (in press). Instrumente zur Lernverlaufsmessung: Gütekriterien und Auswertungsherausforderungen. In M. Hasselhorn, W. Schneider, & U. Trautwein (Eds.), *Formative Leistungsdiagnostik.* Göttingen, Germany: Hogrefe.

Wölfel, U. (1974). *Neunundzwanzig verrückte Geschichten.* Stuttgart, Germany: Thienemann.

Wölfel, U. (2010a). *Achtundzwanzig Lachgeschichten.* Stuttgart, Germany: Thienemann.

Wölfel, U. (2010b). *Siebenundzwanzig Suppengeschichten.* Stuttgart, Germany: Thienemann.