A Systematic Review of Academic Discourse Interventions for School-Aged Children with Language-Related Learning Disabilities

Amy K. Peterson  
*Utah State University*

Carly B. Fox  
*Utah State University*, carly.fox@usu.edu

Megan Israelsen  
*Utah State University*

Follow this and additional works at: [https://digitalcommons.usu.edu/comd_stures](https://digitalcommons.usu.edu/comd_stures)

Part of the *Communication Sciences and Disorders Commons*

**Recommended Citation**

A Systematic Review of Academic Discourse Interventions for School-Aged Children with Language-Related Learning Disabilities

Amy K. Peterson, Carly B. Fox, & Megan Israelsen
Utah State University

Author Note

Amy K. Peterson, Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah; Carly B. Fox, Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah; Megan Israelsen, Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah.

Address correspondence to Amy K. Peterson at Utah State University, 1000 Old Main Hill, Logan, Utah, 84322; amy.peterson@aggiemail.usu.edu; (307) 214-4177

Conflict of Interest

The authors have no competing interests at the time of publication.

Funding

No funding was received for this project.
Abstract

**Purpose:** This systematic review synthesized a set of peer-reviewed studies published between 1985 and 2019 and addressed the effectiveness of existing narrative and expository discourse interventions for late elementary and middle school-aged students with language-related learning disabilities.

**Method:** A methodical search of the literature for interventions targeting expository or narrative discourse structure for students ages 9-14 with group experimental designs identified 33 studies, seven of which met specific criteria to be included in this review.

**Results:** An 8-point critical appraisal scale was applied to analyze the quality of the study design and effect sizes were calculated for six of the seven studies; equivocal to small effects of far transfer outcomes (i.e., generalizability to other settings) and equivocal to moderate near transfer outcomes (i.e., within the treatment setting) were identified. The most effective intervention studies provided explicit instruction of expository texts with visual supports and student-generated learning materials (e.g., notes or graphic organizers) with moderate dosage (i.e., 180-300 minutes across 6-8 weeks) in a one-on-one or paired group setting. Greater intervention effects were also seen in children with reading and/or language disorders, compared to children with overall academic performance difficulties.

**Conclusions:** A number of expository discourse interventions showed promise for student use of learned skills within the treatment setting (i.e., near transfer outcomes), but had limited generalization of skills (i.e., far transfer outcomes).
A Systematic Review of Academic Discourse Interventions for School-Aged Children with Language-Related Learning Disabilities

School-based speech-language pathologists (SLPs) provide services to address a variety of student needs. A significant portion of the SLP caseload includes students with language-related learning disabilities (LLD). Students with specific language disorders and difficulties in reading, writing, and speaking in all levels of academic and social communication (i.e., word, sentence, discourse) are included in LLD. Everyday social interactions with peers and teachers often occur at the discourse level and are difficult for these students. Discourse is any unit of spoken or written communication, longer than one sentence, in any combination of conversational, narrative, persuasive, and expository structures that aid in our interactions with the world (Hughes, LaRae & Schmidek, 1997).

With implementation of the Common Core State Standards Initiative in 2010, students as early as kindergarten are expected to be able to use narrative and expository discourse forms in the classroom (corestandards.org). Narrative discourse includes all storytelling events from early education “share-and-tell” tasks (Temple Adger & Wright, 2015) to advanced productions of complex fictional or personal narratives (Hughes et al., 1997). Narrative discourse requires the use of decontextualized language (i.e., discussing events beyond the immediate context), an important skill for understanding language in classrooms (Bates & MacWhinney, 1979; Curenton & Justice, 2004). Expository discourse is non-narrative, informational language, often presented in academic lessons and textbooks (Nelson, 1993; Hughes et al., 1997); it includes various text structures such as description, explanation, procedure, and persuasion. Ability to use and understand these text structures is critical to academic success.

Comprehension and production of these discourse genres is impacted by poor schema.
retrieval and organization, memory for information within a text, and knowledge of discourse structures (Westby, VanDongen & Maggart, 1989). Understanding the causal framework that underlies narrative discourse (McKinney, Short & Feagans, 1985; Capps, Losh, & Thurber, 2000) and various expository text structures (Dickson, Simmons & Kameenui, 1998; Gersten, Fuchs, Williams & Bakers, 2001) are critical to accessing the curriculum. Students with LLD may gather information from texts in a random fashion, rather than systematically finding and retaining key ideas within an organized mental framework. Inefficient processing strategies may increase strain on cognitive load and negatively impact comprehension and subsequent academic success.

Research indicates that typically developing children tend to have mastered the basic structure of narrative by age 9 (Stein & Glenn, 1982; Merritt & Liles, 1987). Berman and Nir-Sagiv (2007) found that young children are capable of identifying differences in narrative and expository discourse. However, students with LLD often struggle to master production of narrative discourse and may not catch up to their typically developing peers (Snyder & Downey, 1991). To further confound difficulties in comprehension and production for students with LLD, more expository focused texts are included in the curriculum, phasing out the majority of narrative-based lessons around age 9 (Hughes, LaRae & Schmidek, 1997). If the cognitive requisites of narrative discourse (e.g., use of decontextualized language, understanding causal connections, etc.) are not fully developed in students with LLD, attempting expository discourse may prove quite challenging. Pressure to understand the various forms of expository discourse structure without mastery of narrative may even contribute to the “fourth-grade slump” (Chall, Jacobs, & Baldwin, 1991; Merritt & Liles, 1987; Vandewalle et al., 2012).

Studies have shown that at least half of young students with language-related disabilities
have significant literacy and academic issues as they get older (Conti-Ramsden, St. Clair, Pickles, & Durkin, 2012; Dockrell, Lindsay & Connelly, 2009), making it critical for them to receive intervention beyond the early grades. As language-related disabilities are not likely to disappear, the need for specific intervention for late elementary and middle school students is warranted.

Specific interventions, often developed in research settings and tested for efficacy through controlled experiments, are designed by researchers interested in studying certain phenomena or populations. When designing an intervention for students with LLD, it is important for researchers to consider how well their intervention is being implemented in the study and the effect it has on potential participants. In order to measure both of those aspects, fidelity of interventionists and proper qualification of outcome measures must be determined to generate accurate conclusions about the efficacy and effectiveness of an intervention. Studies and definitions of efficacy and effectiveness are done across many fields, including medicine, education, and social sciences like SLP (Kim, 2013; Singal et al., 2014).

Efficacy of an intervention is related how well the treatment was delivered in an ideal circumstance, which supports that the effects of an intervention are due to the intervention itself instead of maturation effects of the population (Singal et al., 2014). Fidelity thresholds of 80% or greater within the controlled environment of the study help to measure the efficacy of studies. While many studies report fidelity statistics, they are not always systematically measured and included as an outcome, indicating that clinicians may need to be wary of results. Additionally, while intervention fidelity checklists may be provided to the interventionists in many studies, those checklists are not always published, making it difficult for clinicians to properly implement those evidence-based practices. This also impacts the replication of studies, leading to
researchers creating new measures for their studies and causes difficulty in establishing valid, reliable measurements (Fixsen, Blase, & Van Dyke, 2019; Olswang & Prelock, 2015).

When studies fail to provide checklists and fidelity information for clinicians to implement an intervention in a real-world setting, they contribute to the ‘research-to-practice gap’. The research-to-practice gap is a gap between the existing literature of research and the real-world practice of clinicians that has been studied for decades in a variety of fields (Olswang & Prelock, 2015; Morris, Wooding, & Grant, 2011). The field of Implementation Science studies the research-to-practice gap, focusing on the effectiveness of innovations, or the ability for high-quality, controlled intervention studies, to generalize to students in real-world settings (Fixsen, et al. 2019; Singal, Higgins, & Waljee, 2014). To address the effectiveness of interventions, this review focuses on near (i.e., within intervention) and far transfer (i.e., generalizability) effects of studies. Near transfer effects, those that are a direct result of the skills learned within treatment and their use within the study context, help to determine efficacy of the study. Far transfer effects are those that generalize to other settings or topics outside of the specific treatment setting or the effectiveness of the treatment overall. Both of these effects are critical to implementing best practices to serve a given population (Singal et al. 2014).

Relevant Prior Reviews

Several reviews have been published synthesizing narrative and expository interventions for school-aged children with and without language disorders. Petersen’s (2011) systematic review of narrative-based interventions for oral narrative macrostructure (i.e., story grammar) and microstructure (i.e., the total number of words, mean length of utterance) in preschool and school-aged children with language or learning disabilities provides valuable information to the field. While Petersen reported low overlap of intervention characteristics across studies, he
SYSTEMATIC REVIEW DISCOURSE

highlighted several major factors contributing to the interventions. All of the studies measured
the development of macrostructure ability through retell and spontaneous generation of
narratives. Several studies also focused on explicit instruction of causality and temporal relations
to develop microstructure skills using picture prompts, narrative illustration tasks, and icons.
Petersen concluded that large effect sizes should be interpreted with caution because of the small
sample sizes and low experimental control across studies.

In a 2016 narrative review, Ward-Lonergan and Duthie summarized interventions
designed to target expository reading comprehension in students with language disorders. The
review does not specify ages or grades; however, the included studies primarily targeted students
in late elementary or beyond. The authors examined a series of interventions that used strategy
approaches (e.g., focus on use of a graphic organizers) or content approaches (e.g., focus on
specific content). Instructors in the content approaches emphasized particular information in the
text through active discussion to help students build mental representations. This review
suggested benefits of content and strategy-based approaches independently and combined for
school-aged children with language disorders; though given the narrative nature of the review no
definitive conclusions can be drawn.

More conclusive results come from Pyle, et al. (2017), a meta-analysis on the effects of
expository text-structure interventions on comprehension in school-aged children and the
moderators of intervention success. The studies included students from grades 2-5 and 8-12 who
were high-achieving, at-risk or learning disabled (LD). The results of this analysis indicated
significant differences between intervention effects, favoring intervention studies where: the
researcher administered the intervention, the length of intervention was 11-20 hours, one or two
text structures were targeted (i.e., cause-and-effect, compare-and-contrast), and participants were
of elementary school age. This meta-analysis provided important information about the efficacy of expository interventions and potential moderators (Pyle et al., 2017). However, the studies included in this review did not target middle-school aged students or students with LLD specifically and 2 studies have been published in this area since 2017.

The current systematic review intended to examine studies that conducted an expository or narrative discourse intervention for students with LLD in late-elementary and middle school grades (i.e., ages 9-14). This review is important to advancing the synthesis of interventions as this population is significant in both narrative and expository discourse development and no other review to date specifically targets this population. Additionally, though fidelity was discussed in the majority of prior reviews, focus on treatment efficacy and effectiveness through near and far transfer outcomes sets this review apart from others. Fidelity information is further analyzed to aid clinicians in knowing which studies have provided their intervention materials to encourage systematic implementation outside of the controlled research environment.

Only group experimental studies that utilized a control or comparison group and specifically measured an intervention were included in this review. Group-level experimental designs were selected to highlight high-quality, rigorous experimental studies that provide evidence-based treatments with potential generalizability to the target population. Group-level experimental designs have the potential to generalize to a broader population due to the homogeneity of the population, random assignment of participants to a control or treatment group in the case of RCTs, and good external validity of results (Institute of Medicine, 2011). In order to further analyze the quality of the included research studies, the Gillam & Gillam (2006) Critical Appraisal Standards were used. These standards provide valuable questions for researchers and clinicians to critically analyze the quality of a research study to encourage use of
research in practice (Appendix A). This review addresses the following questions:

1. What interventions exist that target narrative and/or expository discourse for school-aged children ages 9-14 with LLD?

2. Do interventions conducted with at least 80% fidelity produce both statistically and practically significant improvements in discourse comprehension and/or production?

3. Based on the Gillam & Gillam (2006) Critical Appraisal Standards, at what level of quality were these studies conducted?

Methods

Search Procedure

The Institute of Medicine (2011) guidelines for conducting high-quality systematic reviews were used to guide the procedure for this review. The initial search was conducted using the electronic database for the American Speech and Hearing Association (ASHA), and the EBSCO-hosted databases PsychInfo and ERIC. These databases were selected because of the focus on education, psychology, and speech-language pathology that related to both the population and intervention criteria set by the authors. Search terms were selected based on relevance to the population of interest (i.e., school-aged children with LLD), discourse type (i.e., expository and narrative discourse), modality (i.e., oral and/or written), process (i.e., production or comprehension), and instruction (i.e. intervention, etc.). Related terms were located using the thesaurus tool in PsychInfo to expand search terms for maximum inclusivity. Boolean logic asterisks were used to include all root-word variations (i.e., disab* to include disabled, disability and disabilities) when appropriate systematically combined through chunking similar terms with “OR” and combining terms with “AND”. Systematic searches yielded 500-1000 results per
Additional studies were located through the process of ancestral searching, whereby we examined reference lists of studies included for full-text review for publications that did not populate in our electronic search. An expert in the field of language intervention also suggested several publications as part of a researcher-to-researcher search method (Ukrainetz, T.A., personal communication, March 7, 2019). We excluded: articles that were not published in peer-reviewed journals (i.e., theses and dissertations), book chapters, and studies that were published in languages other than English.

The studies included in this systematic review met the following inclusion criteria:

1. Participants were within the specified age range i.e., (ages 9-14 or grades 4-8). To be included in this review, all of the participants in the study had to be within the specified age range.

2. Participants were students with LLD who had difficulties in language, reading, and writing (Gerber, 1993), qualified for an individualized education plan (IEP), or were considered “at-risk” for a disability by their school or state criteria. The specific disabilities included under LLD are: language disability or impairment, learning disability, and specific language impairment.

3. Interventions targeted narrative or expository discourse. Intervention was defined as a structured activity targeting a student’s production or comprehension in the academic setting directed by a teacher, clinician, or researcher. Interventions also included a measurable outcome on student performance (Cirrin & Gillam, 2008).
4. Studies reported one or more outcome measures of discourse comprehension and/or production. Outcome measures of comprehension included multiple-choice or true/false questions about literal or inferential information from the text and outcome measures of production included written (i.e., essays, short answers) or oral presentations.

5. Studies that employed group-level experimental designs, such as a randomized clinical trial (RCT) or nonrandomized comparison design (i.e., participants are matched across groups) were included in this review. RCTs are considered the highest quality group-experimental designs as they include control groups and higher experimental control than those that include a non-randomized comparison group (Gillam & Gillam, 2006). Group-level experiments produce empirically supported results and have higher external validity for generalization to the broader population (Institute of Medicine, 2011). Quality appraisal standards were employed to further analyze the quality of these studies as recommended by the IOM (Chapter 4).

Methodological Quality

The studies included in a systematic review are rarely conducted by equal standards, thus it was essential to include an indicator of methodological quality to assess implications and risk of bias within studies. The Institute of Medicine recommends a quality analysis step in the systematic review process to reduce the risk of potential bias and provide additional information about implementation (p.178). To address quality in this review, the authors used the Critical Appraisal Standards (Gillam and Gillam, 2006), adapted from Dollaghan (2004). This set of eight questions was used in Cirrin and Gillam’s (2008) systematic review on language interventions for school-aged children with language disorders. Given the similarity of the topic, these standards were appropriate for assessing the quality of studies in this review. The full list
of questions can be found in Appendix A.

The questions included in the *Critical Appraisal Standards* (Gillam & Gillam, 2006) addressed internal validity by assessing: use of a control/comparison group, random assignment of participants, initial group similarities, and assessment blinding. Studies that do not meet these criteria are at risk of making false causal conclusions as confounding variables cannot be ruled out. External validity is addressed through a clearly defined population to increase potential generalizations of treatment effectiveness and reduce over-extension of the results to untested populations. Validity and reliability of measures can impact interpretation of outcomes due to variations in participant-to-participant and administrator-to-administrator results. The final questions addressed statistical and practical significance, which are critical to appraising the intervention effect.

**Coding Procedures**

Records obtained from the initial search were imported to Zotero for organization and categorization by discourse type. In the abstract and title screening, all records were briefly examined for relevance and readily apparent inclusion or exclusion information by one of the three authors and either discarded or considered for full-text review. Records obtained through expert recommendation and ancestral searching were subject to the same screening process. The full-text reviews were coded in REDCap (Harris, Taylor, Thielke, Payne, Gonzalez & Conde, 2009), a secure web-based data management system. The code sheet contained the following subsections: participant information, study setting, intervention characteristics, outcome variable characteristics, and study quality as described above. The coded information was categorized as either a study characteristic or a potential moderating variable. Study characteristics addressed participant age, disability type, discourse type (i.e., expository or narrative), modality (i.e., oral
or written), outcome, and study setting. These variables were considered high-level details of each study design. The moderating variables were selected based on their potential to impact the degree of intervention effectiveness and were identified using previous literature reviews (Pyle et al. 2017; Cirrin & Gillam, 2008). Moderating variables included: group matching procedure, participant diagnosis and areas of language impairment, services provided pre-intervention, length and number of intervention sessions, intervention administrator, and implementation fidelity measurements. In a full-text review, articles were excluded when they failed to meet the inclusion criteria. Following the full-text review, data were extracted through the coding process and the REDCap output. The full coding sheet can be found in supplemental materials.

Objectivity was increased through forced multiple-choice questions to prevent ambiguous coder responses. Before coding, each of the authors reviewed the coding form to clarify any areas of confusion and ensure all critical questions were addressed. All studies were independently double-coded to provide comprehensive inter-coding reliability at a threshold of 80%. Inter-coder reliability was calculated through item-by-item correspondence by dividing the number of items scored in common by the total number of items coded. Across all articles and code sheet subsections, inter-coder reliability ranged from 78.2-94.4%. Coding disagreements were discussed between first and second coders on 100% of the data until a resolution was reached. Discrepancies were subtracted from total number of items and percentage was calculated with an average of 84.6%.

**Effect Sizes**

When sufficient data were provided (i.e., means, standard deviations, number of participants), effect sizes were calculated and characterized as between-group difference or pre-post difference statistics. Effect sizes were included to identify the practical significance of an
intervention and to allow consumers to be more confident that type I or type II errors were not committed. Hedge’s $g$ was selected to measure effect size because it accounts for unequal, small group results better than other effect size measures (Hedges, 1981). Hedge’s $g$ was calculated using a freely available, web-based effect size calculator and interpreted whereby 0.10 is considered small, 0.36 is considered moderate and 0.86 is considered large (Lipsey, Puzio, Yun, Hebert, Steinka-Fry, Cole, Roberts, Anthony & Busick, 2012). This interpretation is preferred over more traditional interpretations, such as Cohen (1977), because it was developed from a systematic review of educational studies and pertains specifically to interventions consisting of a smaller set of targets within a larger curriculum (Lipsey et al., 2012).

Treatment outcomes were identified within the initial coding process and then categorized as either near or far transfer measures by the second author based on the following definitions. After initial categorization, 100% of the measures were reviewed by the first and third authors to determine if any outcome had been mis-categorized. Near transfer outcomes are those that directly relate to what was taught in intervention (e.g. identifying particular structures in an expository text); far transfer outcomes include the application of intervention skills to other contexts (e.g. assessing text structure intervention through reading comprehension on novel texts) (Perkins & Salomon, 1992). Near and far transfer outcomes helped to demonstrate intervention effectiveness. Systematic fidelity measures, those that measure fidelity in various levels of the intervention and directly affect treatment efficacy were determined by the second author and independently agreed upon by the first author. Rather than simply stating if studies utilized a scripted treatment procedure, this review sought to define other fidelity features of the study (i.e., session checklists, observer checklists) to evaluate efficacy and potential for implementation with researcher-provided materials for clinicians.
Results

The original computer search yielded 1,232 records. Of those, 1,202 were excluded based on title and abstract screening for relevance. The ancestral search and researcher-to-researcher recommendations yielded an additional three articles for a total of 33 to be included in full-text review. Based on the full-text review, articles were excluded that: did not include an intervention ($n = 8$), did not specifically target narrative or expository discourse ($n = 1$), were not published in a peer-reviewed journal ($n = 4$), did not have participants within the specified age range ($n = 4$), did not include participants with the specified disability categories ($n = 4$), or were not group-level experimental designs ($n = 5$). The article selection process is detailed in Figure 1. Seven studies met all of the selection criteria (Table 1), summaries of which can be found in Appendix C with related terms as specified by the original authors (e.g., SLI, LLD, etc.).
Figure 1. Flow-chart depicting the full search process.
Effect Sizes

The majority of studies reported significant effects of treatment, with the exception of Griffin and colleagues (1991), who did not find statistically significant effects on immediate or delayed posttest measures (Table 1). Statistical significance is important when analyzing the efficacy of treatment, though to better compare the study effects to one another within this review, Hedge’s g effect sizes were calculated when the appropriate information was provided (Tables 2 & 3). These effect sizes were calculated to further analyze the outcome effects of treatment and to provide clinicians with more evidence of the effectiveness of these treatments. The majority of studies had small effect sizes with two studies having effects in the moderate-large range. Scanlon (1996) had large treatment effects for participants’ creation of graphic organizers in both TD and LLD groups post-treatment. Ukrainetz (2019) had large effects on quality of notes in “quick and easy” and “bulleted/picto” categories favoring the intervention group. Outcomes measuring use of full, open/close sentences, and modified sentences were also moderate-large for Ukrainetz (2019).
### Table 1. Study Characteristics – participants, design, interventions and outcomes

<table>
<thead>
<tr>
<th>Citation</th>
<th>Participants</th>
<th>Study Design</th>
<th>Intervention Type</th>
<th>Intervention Target</th>
<th>Outcome Measure</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnine, D., &amp; Kinder, D. (1985)</td>
<td>N=27 LLD or at-risk; unequal group sizes (14 generative, 13 schema) performance-level matched</td>
<td>Random assignment with comparison group</td>
<td>Generative expository discourse intervention (treatment condition) and narrative discourse (comparison condition) adopted from 3-4 grade level texts</td>
<td>1. Reading comprehension 2. Oral production</td>
<td>Rubric-based, measured pre- and posttest with follow-up</td>
<td>YES $p &lt; .001$ for pre-test to transfer-test comparisons of narrative and expository comprehension and narrative retell indicating a training effect; high correlation of performance between transfer and maintenance tests</td>
</tr>
<tr>
<td>DiCecco, V. M., &amp; Gleason, M. M. (2002)</td>
<td>N = 24 with LLD; equal participants, performance-level matched</td>
<td>Cluster randomization by school with comparison group</td>
<td>Graphic organizer intervention with curricular expository text</td>
<td>1. Expository reading and listening comprehension Oral and written production</td>
<td>Overall performance rating, pre- and posttest</td>
<td>YES $p &lt; .001$ for time of test on written measures and content knowledge for both groups; $p = .0007$ for treatment group on inclusion of relational statements at post-test</td>
</tr>
<tr>
<td>Study</td>
<td>N</td>
<td>Group Characteristics</td>
<td>Intervention Details</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Griffin, C. C., Simmons, D. C., & Kameenui, E. J. (1991)** | N = 28 with LLD or at-risk; equal participants, performance-level matched | Random assignment with comparison group                                               | Graphic organizer intervention with curricular expository text                      | 1. Reading comprehension  
2. Oral and written production                                                      | Rubric-based, measured pre- and posttest with follow-up                          |
|                               |                    |                                                                                       |                                                                                       | **NO** p > .05 for both groups at immediate and delayed posttest on oral retell, production, and choice-response measures |
| **Hebert, M., Bohaty, J., Nelson, J. R., Roehling, J., & Christensen, K. (2018)** | N = 12, with LD or at-risk; unequal group sizes (7 experimental, 5 control), performance-level matched | Random assignment with control group                                                  | Note-taking and text structure intervention with researcher-developed expository text | 1. Expository reading comprehension  
2. Written production                                                                  | Check-list, measured pre- and posttest                                          |
|                               |                    |                                                                                       |                                                                                       | **YES** p < .05 for structures, compare/contrast, and sequence writing proximal outcome measures; p = .007 for structures identification distal outcome measure |
| **Starling, J., Munro, N., Togher, L., & Arciuli, J. (2012)** | N = 43 with SLI; unequal group sizes (22 treatment, 21 control); matching unclear between groups | Random assignment with control group                                                | Explicit expository language instruction for classroom teachers                    | 1. Reading and listening comprehension  
2. Oral and written production                                                         | Norm-referenced, measured pre- and posttest with follow-up                       |
|                               |                    |                                                                                       |                                                                                       | **YES** p < .05 for Listening Comprehension and Written Expression  
*WIAT-II* scores at post-test with students in treatment group outperforming control |
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Characteristics</th>
<th>Assignment Method</th>
<th>Intervention</th>
<th>Measures</th>
<th>Conclusion</th>
</tr>
</thead>
</table>
| Scanlon, D. (1996) | N = 204, both TD and LD; unequal group sizes (109 treatment, 95 control), closely matched based on chronological age | Non-random assignment with comparison group; classrooms selected by teachers | ORDER strategy intervention with researcher-developed expository text | 1. Expository reading comprehension 2. Written Production | Overall performance rating, measured pre- and posttest  
*YES* \( p < .0005 \) for creating graphic organizers, favoring treatment over comparison at post-test |
| Ukrainetz, T. A. (2019) | N = 44, with LLD or SLI; equal group sizes, performance-level matched | Random assignment with control group | Sketch and Speak intervention with researcher-developed expository text | 1. Expository reading comprehension 2. Oral and written production | Rubric-based, measured pre- and posttest  
*YES* \( p = .001 \) for *Quick* and *Bullet/Picto* categories and quality of notes at post-test, favoring the treatment group |
Table 2. Near-Transfer Measure Effect Sizes

<table>
<thead>
<tr>
<th>Study</th>
<th>Name(s) of Measure</th>
<th>Medium</th>
<th>Description</th>
<th>Hedge’s g Between Group Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnine, D., &amp; Kinder, D.</td>
<td>Information units recalled (IUR)</td>
<td>Oral</td>
<td>Generative group instructors identified key chunks of information and discussed significance to the text. At post-test the mean number of information units recalled during expository retell were assessed.</td>
<td>IUR: $g = 0.175$</td>
</tr>
<tr>
<td>(1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DiCecco, V. M., &amp; Gleason, M.</td>
<td>Number of words written for all essays (TNW)</td>
<td>Written</td>
<td>Number of words written was calculated for each written measure to assess general writing ability.</td>
<td>TNW: $g = 0.114$</td>
</tr>
<tr>
<td>M. (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griffin, C. C., Simmons,</td>
<td>Total units recalled Total score Production &amp; Choice</td>
<td>Oral</td>
<td>Total units recalled assessed by the number of key details identified in the oral report retell of the intervention text. Total score accounted for the importance of the unit recalled. Production and choice responses assessed comprehension of the intervention text.</td>
<td>TUR: $g = 0.488$</td>
</tr>
<tr>
<td>D. C., &amp; Kameenui, E. J.</td>
<td></td>
<td></td>
<td></td>
<td>TS: $g = 0.418$</td>
</tr>
<tr>
<td>(1991)</td>
<td></td>
<td></td>
<td></td>
<td>P: $g = 0.573$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C: $g = 0.539$</td>
</tr>
<tr>
<td>Hebert, M., Bohaty, J.,</td>
<td>Structure-identification Structure notes</td>
<td>Written</td>
<td>The structure identification measure assessed ability to choose the correct type of expository structure when reading assessment text. Students were instructed to take notes on additional passages and notes were scored for the number of structures present.</td>
<td>S-ID: $g = 0.73$</td>
</tr>
<tr>
<td>Nelson, J. R., Roehling, J.,</td>
<td></td>
<td></td>
<td></td>
<td>SN: $g = 0.906$</td>
</tr>
<tr>
<td>&amp; Christensen, K. (2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scanlon, D. (1996)</td>
<td>Creation of graphic organizer</td>
<td>Written</td>
<td>Ability to design a graphic organizer with critical information from an expository passage was assessed. Typically developing (TD) and those with LD were assessed separately for each group.</td>
<td>TD: $g = 1.08$</td>
</tr>
<tr>
<td>Ukrainetz, T. A. (2019)</td>
<td>Notes quantity &amp; quality</td>
<td>Written</td>
<td>Note quantity measured by the number of notes; quality measured through five indices: format (2), brief, sufficient and paraphrasing.</td>
<td>Quan: $g = 0.163$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Qual: $g = 1.199$</td>
</tr>
</tbody>
</table>

Note: The effect sizes are for between group differences (i.e. control/comparison group versus intervention group).
Table 3. Far-Transfer Measure Effect Sizes

<table>
<thead>
<tr>
<th>Study</th>
<th>Name(s) of Measure</th>
<th>Modality</th>
<th>Description</th>
<th>Hedge’s $g$ Between Group Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnine, D., &amp; Kinder, D. (1985)</td>
<td>Expository transfer items, Inferential comprehension, Literal comprehension</td>
<td>Written</td>
<td>ETI = measure of expository knowledge transfer. Participants were administered a comprehension test that required application of learned skills including three literal and two inferential questions.</td>
<td>ETI: $g = 0.427$ Inferential comp: $g = 0.205$ Literal comp: $g = 0.281$</td>
</tr>
<tr>
<td>DiCecco, V. M., &amp; Gleason, M. M. (2002)</td>
<td>Content Knowledge Fact Quiz</td>
<td>Reading comprehension</td>
<td>A series of 8 fact quizzes were administered throughout the intervention to assess comprehension of factual information in the text. Only the final quiz effect size is presented here.</td>
<td>Fact quiz: $g = 0.073$</td>
</tr>
<tr>
<td>Griffin, C. C., Simmons, D. C., &amp; Kameenui, E. J. (1991)</td>
<td>None</td>
<td>N/A</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Hebert, M., et al. (2018)</td>
<td>Structure comprehension</td>
<td>Written &amp; Reading comprehension</td>
<td>Students answered a series of 20 comprehension questions related to a previously taught expository text.</td>
<td>Comprehension: $g = 0.006$</td>
</tr>
<tr>
<td>Scanlon, D. (1996)</td>
<td>None</td>
<td>N/A</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Ukrainetz, T. A. (2019)</td>
<td>Full sentences Open/Close Sentences Modified sentences</td>
<td>Written</td>
<td>Measures of quality for the posttest written reports on expository texts included: use of full sentences, opening/closing statements and modifications to sentences presented in original text.</td>
<td>Full: $g = 0.480$ O/C Sentences: $g = 0.492$ Modified Sentences: $g = 0.640$</td>
</tr>
</tbody>
</table>

Note. The effect sizes listed in this table pertain to between group differences (i.e., control/comparison group versus intervention group). No effect sizes are reported for Griffin et al. (1991) or Scanlon (1996) as they could not be calculated.
Near and Far Transfer Outcomes

Coded information on transferability revealed that the majority of outcomes across studies were near transfer (i.e., posttest measured ability to do what was directly taught in intervention). See Table 2 for description of near transfer effects and Table 3 for far transfer effects by study. Two of the studies only measured near transfer tasks (Griffin et al., 1991; Scanlon, 1996). The studies that included both near and far transfer tasks had mixed effect sizes. Carnine and Kinder (1985) had larger effect sizes for their far transfer tasks related to inferential and literal comprehension (Table 3), compared to their near transfer task on recalling information units (Table 2). Alternatively, Hebert et al. (2018) had moderate to large effect sizes on experimenter-designed near transfer tasks of note taking and structure identification (Table 2), but had an equivocal effect size on a far transfer comprehension task (Table 3). In general, far transfer tasks are more indicative of generalization and near transfer tasks do not necessarily extend beyond the intervention setting.

Fidelity

Of the seven studies included in this review, four of the studies reported fidelity (Table 4). Of these studies, only three reported 80% or higher fidelity, an important consideration for treatment effects due to intervention rather than maturational effects of participants. Additionally, only two of the studies that reported fidelity supplied materials or checklists for clinicians to implement the intervention in practice (Ukrainetz, 2019; Hebert et al., 2018).
Table 4. Fidelity of Intervention Implementation

<table>
<thead>
<tr>
<th>Study Authors</th>
<th>Systematic Intervention Administration</th>
<th>Checklists Provided in Publication</th>
<th>Fidelity Tracked</th>
<th>Fidelity Level</th>
<th>Fidelity Above 80%?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnine, D., &amp; Kinder, D. (1985)</td>
<td>Scripted lessons</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>DiCecco, V. M., &amp; Gleason, M. M. (2002)</td>
<td>Scripted lessons, Observer checklists</td>
<td>No</td>
<td>Yes</td>
<td>Not reported</td>
<td>N/A</td>
</tr>
<tr>
<td>Griffin, C. C., Simmons, D. C., &amp; Kameenui, E. J. (1991)</td>
<td>Scripted lessons with expectation of teacher to add information</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Scanlon, D. (1996)</td>
<td>Example scripts for lessons, Session behavior checklist, Observer checklist</td>
<td>Not provided</td>
<td>Yes</td>
<td>21.9% (averaged across multiple measures)</td>
<td>No</td>
</tr>
<tr>
<td>Ukrainetz, T. A. (2019)</td>
<td>Scripted introduction of new skills, Treatment fidelity checklists for interventionist and observer</td>
<td>Yes</td>
<td>Yes</td>
<td>95%</td>
<td>Yes</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>Appendix B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quality of Intervention Studies

The Critical Appraisal Standards (Gillam & Gillam, 2006) questions are found in Appendix A. Appraisal points for each study based on this evaluation of quality indicate that clinicians can have moderate confidence in the results presented in this synthesis (Table 5). All seven of the included studies utilized a comparison/control group within their experiment. The majority of studies also earned points for statistical and practical significance of treatment and for randomly assigning participants. Random assignment of participants within an experiment is ideal to provide the greatest experimental control, reduce potential maturation effects, and increase potential for generalization (Gillam & Gillam, 2006). No-treatment control groups, like the business-as-usual control used in Ukrainetz (2019), or alternate treatment groups, like in Hebert et al. (2018) can be used to examine intervention effects. Alternate treatment groups are used as a more pragmatic way to provide some form of intervention to all when a no-treatment option is impractical. For example, in Hebert et al. (2018) an alternate treatment was used because intervention occurred after school. Scanlon (1996) is the only study in this review that did not use random assignment of participants to the experimental groups, therefore not earning the Critical Appraisal Standards (Gillam & Gillam, 2016) point for this item.

Many studies lost points for not describing the blinding procedures and for not using previously established, valid and reliable outcome measures. Only one of the studies met the “measures” appraisal value for using previously established, valid and reliable measures for outcomes with Starling et al. (2012) using standardized tests at pre/post and follow-up testing sessions. The highest quality interventions included in this review by these standards were Ukrainetz (2019) and Starling et al. (2012), both earning 7 of 8 quality points. Ukrainetz employed a business-as-usual control group, included participants with LLD who had reading
and decoding impairments, and provided six intervention sessions, 30 minutes in length in a one-on-one or paired group context, administered by a trained SLP (2019). SLPs in the Ukrainetz (2019) study trained students on note-taking from expository texts combined with verbal rehearsal of complete sentences to increase student comprehension of grade-level material through a variety of ‘real student life’ discourse tasks. Starling and colleagues implemented a collaborative treatment where an SLP trained classroom teachers on increasing the impact of specific language instruction through: breaking down large amounts of information into smaller, visually distinct sections, using picture supports, providing descriptions for new vocabulary, and placing questions on the same page as the text in small group or one-on-one discussions over a 10-week period (2012).
Table 5. Gillam & Gillam (2006) Critical Appraisal Standards points for included articles

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group or treatment comparisons</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Random assignment</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Participant information</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Initial group similarity</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
<td>No</td>
<td>Yes**</td>
<td>Yes*</td>
</tr>
<tr>
<td>Blinding</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
</tr>
<tr>
<td>Measures</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes***</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Practical significance</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Total appraisal points</td>
<td>6/8</td>
<td>5/8</td>
<td>6/8</td>
<td>6/8</td>
<td>7/8</td>
<td>5/8</td>
<td>7/8</td>
</tr>
</tbody>
</table>

Potential Moderating variables

No conclusive statements about the impact of moderating variables can be made given the nature of this review, however, there were a number of consistent factors across the study designs and implementation that may have affected outcomes that warrant future investigation. Moderating variables were determined based on those identified in previous reviews (Pyle et al., 2017; Cirrin & Gillam, 2008). First, studies with participants diagnosed with only one component of LLD (i.e., a singular impairment in either reading or language) showed higher posttest gains than those that included participants with overall academic performance difficulties (i.e., participants in Carnine & Kinder, 1985). This may be due to a number of factors, including the intervention itself, the group size and matching procedures, or the population of students. A future meta-analysis might therefore consider evaluating the number/type of impairment on responsiveness to intervention.

A second potential moderating variable identified was intervention dosage. The interventions ranged in duration from 4 days to 19 weeks and in instructional time from 180 to 570 minutes. We found larger effect sizes for the interventions with a moderate dosage (i.e., 180-300 minutes across 6-8 weeks). Previous reviews focused on reading interventions have found larger effect sizes during shorter interventions for students with learning disabilities (Elbaum et al., 2000; Scruggs et al., 2010; Pyle et al., 2017). Our analyses determined that Ukrainetz (2019) had the largest effect size and the highest quality scale rating in only 6 sessions of intervention while one of the smallest intervention effect sizes came from a study that had only four sessions (Griffin, 1991). Additionally, Ukrainetz (2019) found statistically significant results for outcome measures while Griffin and colleagues (1991) did not. Differences in intervention length, timing, and setting, make it beyond the scope of this review to discern how the length of treatment
impacts the effect sizes found for this review without replication of these high-quality studies.

A final potential moderating variable was the intervention administrator. The Ukrainetz (2019) study showed the greatest posttest gains with intervention administered by familiar, trained SLPs in a one-on-one or one-on-two setting. This was followed by whole-class interventions taught by the classroom teacher (Scanlon et al., 1996; Starling et al., 2012), then pairs or individual participants instructed by a research assistant (Hebert et al., 2018), and finally interventions administered by resource teachers instructing groups of participants (DiCecco & Gleason, 2002; Carnine & Kinder, 1985; Griffin et al., 1991). Again, without conducting a meta-analysis, it is impossible to conclude whether any of these factors moderated intervention effects, warranting future investigation of these variables and interventions.

**Discussion**

This review aimed to identify current interventions for children ages 9-14 with LLD for expository and/or narrative discourse, the quality of the study designs, the fidelity of the experimental treatment, and the statistical and practical significance of the intervention. The synthesis included seven interventions, which were primarily strategy approaches (i.e. graphic organizers, note-taking strategies) targeting expository discourse. Six of the seven studies reported statistically significant posttest differences between intervention and control/comparison groups on at least one outcome measure. We calculated Hedge’s $g$ for 6 of 7 studies based on the descriptive statistics provided by the original authors, and found effect sizes that ranged from small to large, with the Ukrainetz (2019) *Sketch and Speak* intervention producing the largest group difference ($g = 1.199$) at post-test on near transfer tasks. This supports the potential for explicit instruction in expository discourse in the area of written note quality within the treatment setting. Overall this review found a small number of empirically supported interventions
analyzed at the group experimental level for this population. The majority produced some level of significant change in near transfer skills, however, far transfer effects were primarily small. At the outset of this synthesis, we expected to find both narrative and expository interventions because of the continued difficulty with narrative discourse in students with LLD (Snyder & Downey, 1991). Despite the increased use of expository text structure in the later grades, we anticipated continued treatment in both patterns of discourse for students with LLD because of the potential for continued difficulty with narrative structure (Stein & Glenn, 1982; Merritt & Liles, 1987). We found only one study that examined a narrative intervention for this population while also targeting expository language skills (Carnine & Kinder, 1985); all other studies measured outcomes in expository discourse structures. The low number of group level experimental design studies on narrative intervention for this population could identify a need for future research in this area.

**Efficacy and Effectiveness of Interventions**

To analyze the efficacy of these studies, an 80% threshold of fidelity was set by the authors. Fidelity of implementation within studies is important for clinicians to be able to determine that the effects of treatment were due to the intervention itself. Three studies met this threshold, but only two of the studies provided access to fidelity checklists for clinicians to better implement the innovation in practice (Ukrainetz, 2019; Hebert et al. 2018). The small number of studies providing materials to clinicians may increase the research-to-practice gap and result in limited effectiveness of interventions in real-world situations. In order for clinicians to appropriately implement these interventions, an increased report of fidelity and more accessible treatment materials is necessary in future studies to increase clinician implementation in real-world settings. Appropriate implementation of an intervention depends on adequate researcher-
to-clinician communication, investigation of core components of treatment, observations of implementation in real-world situations, and active studies of research in practice through implementation teams (Fixsen et al. 2019). Researchers can better address the research-to-practice gap by providing examples, fidelity checklists, and materials upon request to clinicians.

**Generalizability**

This review generalizes to populations well-known by SLPs as students with LLD make up significant portions of the caseload in a school setting from kindergarten to 12th grade. The results of this review suggest that interventions for expository discourse, including instruction with graphic organizers to highlight connections between main points and strategies like note-taking with verbal rehearsal (Ukrainetz, 2019), can be beneficial for increasing student understanding. Highly structured tasks with explicit instruction (i.e., ORDER, *Sketch and Speak*) and less structured note-taking tasks used in Hebert et al. (2018) and DiCecco & Gleason (2002) both benefitted students with LLD.

Though all of the intervention strategies were explicitly taught, there was a great deal of variability in the validity of the measurements used and the amount of student carryover after intervention. The amount of student progress varied significantly based on the level of explicit instruction and scaffolding provided to the students. Whole class instruction provided the best outcomes (i.e., largest effect sizes) for students second only to one-on-one instruction (Ukrainetz, 2019). The research designs included in this review were all experimental, high-quality intervention studies. Some of the studies included a randomly assigned alternate treatment group instead of a control group (Carnine & Kinder, 1985; Hebert et al., 2018) to provide some form of intervention to all study participants. The main caveat to alternate treatment groups, however, is that it becomes difficult to disentangle which posttest effects are due to the intended components
of the experimental intervention, and which may be due to components common to both methods of instruction.

**Limitations**

The usefulness of this systematic review is impacted by the strength of the research designs and by factors related to publication bias. There were only 7 studies that met the criteria set by the authors, limiting the possible implications of this review on intervention for the LLD population. The validity and reliability of measures across treatments is also a limitation as the majority of studies in this synthesis used self-developed tools to analyze performance. Only one study used previously established measures by using the WIAT subtests at pre-post and follow-up testing of student performance (Starling et al. 2012). Though standardized tests are valid and reliable, they are not designed to measure improved performance over short periods of time in most cases. Additionally, the use of different tools across studies made it difficult to compare intervention effectiveness and impacted the generalizability of treatments. With replication and validation of these measures through follow-up research, more informed treatment methods for expository discourse may be available for clinicians in the future. Incomplete reporting of descriptive statistics across studies is another potential bias that could have affected the interpretation of effect sizes. Given that some studies reported incomplete descriptive statistics, effect sizes could not be calculated for all outcome measures. It is possible that some authors might have only reported the descriptive statistics for their most significant results; however, without complete information effect sizes are unable to be determined.

Publication biases may have also impacted the availability of studies to be included in this review. Studies with negative or equivocal results are often not published, limiting the availability of research, though the degree of bias is difficult to measure. Case studies and single-
subject designs are standard in speech pathology and educational research, though we decided not to include them in this review due to their limited generalizability and potential for biases (Institute of Medicine, 2011). Of the 33 studies examined for full-text review there were 4 studies excluded because of study design (e.g. quasi-experimental or case study) and 6 excluded for not providing intervention specific to students with LLD. The limited availability of studies with robust, group-level research designs targeting students with LLD also impacted the results of this synthesis, though the Critical Appraisal Standards (Gillam and Gillam, 2006) helped to provide information about the quality of implementation across studies.

When looking for empirically supported treatments beyond what are listed in this review, clinicians may need to consider adapting interventions developed for alternate populations (e.g. younger children or non-LLD), or consider interventions conducted at good, but lower levels of experimental rigor, such as single-subject designs. Going forward, more high-quality studies in the specific area of speech-language pathology would significantly increase the quality of practices available to SLPs.

Implications for Clinicians and Clinical Researchers

Clinicians and researchers should interpret the effect sizes reported in this review with caution. All students, regardless of ability, benefitted from explicit instruction with expository texts, though the far transfer of skills into other contexts was small if available. In the majority of studies included in this review, students demonstrated learning of specific strategies taught during the intervention (i.e., creating graphic organizers or identifying text structures) but there was minimal evidence that these skills generalized to comprehension or production measures at post-testing. No study in this review used delayed follow-up testing on student independence and use of strategies in other learning environments. Though the use of graphic organizers can help
students with LLD to have more understanding of the connections within and across texts,

explicit instruction of skills beyond this is likely necessary to increase student benefit. The use of explicit strategy instruction within expository texts shows promise for improved comprehension of discourse for students with LLD. Additional work is needed to determine the best method for teaching far transfer of skills to unlearned contexts, which continues to prove difficult for children with LLD.

We have determined a gap in the research of narrative discourse interventions for continued development and maintenance in older students with LLD that may be addressed by future research. Research with younger students has demonstrated that weakness with narrative discourse significantly impacts academic performance in students with disabilities (Bloome, Katz, & Champion, 2003; Stein & Glenn, 1982). Evidence further suggests that narrative abilities do not spontaneously develop over time for students with language impairments (Snyder & Downey, 1991). Therefore, narrative intervention should not cease for older students, especially if the understanding of this discourse structure is not mastered. Based on this review, studies of narrative interventions in students over age 9 would improve empirically supported treatments available for clinicians.

Finally, there is much research on the use of graphic organizers within intervention for expository text structures, but it is unclear how comprehension is measured across studies. It is common to use written and spoken output to determine comprehension of materials (i.e., CCSS 4th grade standards), though it is unclear which interventions best impact student performance with discourse level information. Ukrainetz (2019) found increased student performance on oral reports and comprehension of intervention materials, but these effects did not transfer to the comprehension tests at post-testing with new material. Areas of future research include
examining the generalizability of interventions from research to practice and determining whether positive intervention effects are sustained through delayed follow-up. An increased research focus on the outcome of global student performance and maintenance could increase the potential of interventions for SLPs and success of students with LLD.

Acknowledgements

The authors would like to thank Ronald B. Gillam for his guidance and encouragement on this project and the revision process. We would also like to thank Teresa A. Ukrainetz for acting as a “researcher-to-researcher” search method to strengthen the reaches of this review.
References


Common Core State Standards Initiative (n.d.). Retrieved from:
http://www.corestandards.org/ELA-Literacy/RI/K/


Doi:10.3102/0034654308325998


SYSTEMATIC REVIEW DISCOURSE

Figure Legends:

Figure 1. Flow-chart depicting the full search process.

Table 1. Study Characteristics – participants, design, interventions and outcomes

Table 2. Near-Transfer Measure Effect Sizes

Note. The effect sizes are for between group differences (i.e. control/comparison group versus intervention group).

Table 3. Far-Transfer Measure Effect Sizes

Note. The effect sizes listed in this table pertain to between group differences (i.e. control/comparison group versus intervention group). No effect sizes are reported for Griffin et al. (1991) or Scanlon (1996) as they could not be calculated.

Table 4. Fidelity of Intervention Implementation

Table 5. Gillam & Gillam (2006) Critical Appraisal Standards points for included articles


Appendix A


Appendix B

Table of Search Terms Used in PsychINFO

Appendix C

Summary of included studies using author’s terminology, organized by intervention setting
## Appendix A

Gillam & Gillam (2006) *Critical Appraisal Standards*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison Group</td>
<td>Was there a control group and at least one or more treatment groups within the study?</td>
</tr>
<tr>
<td>Random Assignment</td>
<td>Was random assignment used to assign participants to control or treatment groups?</td>
</tr>
<tr>
<td>Participants</td>
<td>Was enough information provided about participants (i.e., age, gender, ethnicity, SES, speech and language abilities, and/or cognitive status) within the study?</td>
</tr>
<tr>
<td>Initial Group Similarity</td>
<td>Before treatment, were the groups similar on all important ways (e.g., age, ability level, etc.)?</td>
</tr>
<tr>
<td>Blinding</td>
<td>Were the people who administered and scored the assessments blind to which groups the participants were placed in?</td>
</tr>
<tr>
<td>Measures</td>
<td>Were the measures (both formal and informal) used to obtain outcomes both valid and reliable?</td>
</tr>
<tr>
<td>Statistical Significance</td>
<td>Were $p$-values reported that were less than 0.05?</td>
</tr>
<tr>
<td>Practical Significance</td>
<td>Were moderately-large $\eta^2$ values or standardized $d$ measures reported? If not, can these values be calculated from the data included?</td>
</tr>
</tbody>
</table>

### Appendix B

Table of Search Terms Used in PsychINFO

<table>
<thead>
<tr>
<th>Chunk</th>
<th>Term Type</th>
<th>String of Search Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Discourse</td>
<td>Express* OR oral OR written OR production OR spoken OR “oral communication” OR “oral reading” OR “written communication” OR “written language” OR discourse OR text</td>
</tr>
<tr>
<td>S2</td>
<td>Discourse</td>
<td>Receptive OR comprehension OR understanding OR “comprehension test”</td>
</tr>
<tr>
<td>S3</td>
<td>Discourse</td>
<td>Narrative* OR expository OR information OR stories OR story OR essay* OR storytelling OR “essay testing”</td>
</tr>
<tr>
<td>S4</td>
<td></td>
<td>S1 AND S2 AND S3</td>
</tr>
<tr>
<td>S5</td>
<td>Intervention</td>
<td>Teach* OR instruct* OR intervention* OR treatment* OR framework* OR educat* OR “teaching method” OR “individualized instruction” OR “programmed instruction” OR therapy OR “treatment outcome”</td>
</tr>
<tr>
<td>S6</td>
<td></td>
<td>S4 AND S5</td>
</tr>
<tr>
<td>S7</td>
<td>Participant</td>
<td>Child* OR “school-age” OR elementary OR “elementary education” OR “elementary school student” OR “middle school” OR “junior high” OR “middle school education” OR “middle school student” OR “junior high school student”</td>
</tr>
<tr>
<td>S8</td>
<td>Participant</td>
<td>Impairment* OR disabili* OR disorder* OR delay* OR disabled OR disadvantaged OR “delayed speech” OR “delayed development”</td>
</tr>
<tr>
<td>S9</td>
<td></td>
<td>S7 AND S8</td>
</tr>
<tr>
<td>S10</td>
<td></td>
<td>S4 AND S5 AND S9</td>
</tr>
</tbody>
</table>
Appendix C

Summary of included studies using author’s terminology, organized by intervention setting

Classroom-Based Interventions

Scanlon et al. (1996) taught students to organize essential information into a graphic organizer to target reading comprehension and written production of expository texts. The researchers used a non-randomized quasi-experimental design where whole classrooms were assigned to treatment (109 students) or control (95 students) groups. Classrooms in both groups included both typically developing (TD) students and those with learning disabilities (LD). Students within the experimental group were taught to use the ORDER strategy which involved five steps: 1) open your mind & take notes, 2) recognize the structure, 3) design an organizer, 4) explain it, and 5) recycle it. The students were also taught four major expository text structures including sequence, compare/contrast, descriptive, and problem-solution. FLOW was a sub-strategy within step 3 that helped students create a graphic organizer and included: (1) finding and listing important information, (2) looking and checking for appropriate text structure, (3) organizing the information using numbers or symbols, and (4) working out an organizer to create the final visual product. These strategies were taught by familiar teachers in 4, 25-minute lessons using expository texts from their general curriculum in history or social studies. Originally, teachers agreed to teach one 25-minute session per week for the remainder of the school year, though this was not completed by most of the teachers and likely impacted student performance at post-test. Several materials were included to facilitate proper instruction of the intervention at the classroom level, including an instruction manual containing descriptions of both the FLOW and ORDER strategies, practice activities, transparent sheets for overhead projection, posters for the classroom, and graphic organizer checklists. The authors reported statistically significant
differences favoring the experimental group on the creation of graphic organizers (p < 0.005) with no interaction (p > 0.05) between condition and group, TD or LD. They proposed that this was an effective strategy to improve comprehension of expository text, especially for students with LD, if provided with explicit instruction.

Starling et al. (2012) conducted a classroom-based study of a collaboration intervention for speech-language pathologists (SLPs) and general education classroom teachers in two secondary-education schools randomly assigned to treatment or control conditions. The goal of the intervention was to: increase teachers’ use of explicit instructions, increase repetition and rephrasing of information, and allow for increased student processing time. Teachers in the intervention condition participated in a training program led by the SLP once a week for 50-minutes over 10 weeks. SLPs taught teachers to break down large amounts of information into smaller, visually distinct sections, use picture supports, provide descriptions for new vocabulary, and place questions on the same page as the text in small group or one-on-one discussions. The SLP also observed at least three class periods where the teacher implemented the new strategies. Visual planners and outlines of the task sequence were used to aid in student production and teachers provided direct vocabulary instruction based on a three-tier vocabulary system. Performance measures were conducted for experimental group teachers at pre, post, and follow-up using a structured face-to-face interview with the Levels of Use tool (i.e., LoU) adapted from the Concerns-Based Adoption Model (CBAM; Hord, Rutherford, Huling-Austin, & Hall, 2006) to measure change in 7 distinct areas of instruction for each teacher. A total of 43 students with language impairment (21 treatment, 22 control) were given standardized spoken and written examinations at pre, post, and follow-up to measure intervention outcomes. Students within the treatment classrooms made significant improvements on written expression (p = 0.02) and
listening comprehension (p = 0.033) as measured by subtests of the Wechsler Individual Achievement Test, Second Edition, Australian Standardised Edition (WIAT-II; Wechsler, 2007).

There were no significant changes on the oral expression (p = 0.429) and reading comprehension subtests (p = 0.833) compared to the control students.

**Small Group Interventions**

DiCecco & Gleason (2002) taught students with LDs to use graphic organizers (GO) within a common grade-level social studies textbook. The students ranged in age from sixth to eighth grade (mean age of 13.5). The study included 24 participants randomly assigned to either the GO (n = 12) or no-GO (n = 12) condition. Participants were further assigned to small groups of 4. The intervention was conducted in resource rooms by trained special education teachers with varying levels of professional experience. Students in the experimental groups were instructed during a typical 40-minute class period each day for a total of 20 school days. Instruction included explicit, visual and verbal representations of relationships and details (i.e., instruction on graphic organizers displayed for the group) from the texts while the teachers read the texts aloud. The no-GO group was performance matched and given the same, scripted instruction and explicit verbal review, but did not receive a visual representation of details during the review session. Outcomes were measured with content knowledge multiple-choice tests administered pre and post-treatment, content quizzes throughout intervention, and two domain knowledge essays. The results indicated a statistically significant posttest difference on the number of relational knowledge statements in written essays (p < 0.005) with the GO group outperforming the no-GO group. Meaning that participants who received intervention with the GO made more connections within the text than those who were not trained on GOs. The authors hypothesized that the quizzes and tests may not have been equivalent, citing better student
Carnine & Kinder (1985) compared schema-based (n=13) and generative (n=14) teaching methods for increasing comprehension of expository and narrative texts in grades 4-6. Students were referred by teachers to the study because of comprehension difficulties and ranged in performance from TD to “mildly handicapped” based on district qualification. Experienced teachers provided the intervention in 20-30 minute small group sessions (i.e., 3-5 students) 3-4 times per week for a total of 32 sessions. There were nine expository text lessons in which students read one text aloud together sentence-by-sentence and 10 narrative lessons with three texts: one teacher read-aloud, one student read-aloud, and one read silently each session. Outcome measures were based on comprehension test performance. The authors predicted that the schema group would outperform the generative group in both narrative comprehension and expository maintenance on transfer items based on previous studies, however no statistically significant group differences were found, $p > 0.05$.

Griffin et al. (1991) taught late elementary school-aged students with learning disabilities to use graphic organizers (i.e., GOs) to improve reading comprehension and recall of information from scientific expository texts. The study included 28 participants with identified LD, matched by performance level, and randomly assigned to either GO intervention or comparison groups. Two experienced special education teachers acted as the treatment administrators. Participants were taught in 4 consecutive, 45-minute sessions on a text about fossil fuels, regardless of condition. The students in the GO condition were provided with visual aids to highlight and explain relationships between critical facts from the text, while the no-GO group was given a bulleted list of the same facts. Outcomes were measured through oral retells, written response
items (i.e., fill in the blank or short answer), and multiple-choice questions on the learned material. The authors reported non-significant post group differences between GO and no-GO conditions, p > 0.05.

**One-on-One or Paired Intervention**

Hebert et al. (2018) taught late elementary school-aged children with LD note-taking and text-structure identification strategies to improve expository text comprehension and subsequent written production. Twelve participants were randomly assigned to the expository text intervention (n =7) or to an alternate treatment group targeting narrative discourse (n =5) with one-on-one or paired group instruction. Expository text intervention was administered through two lesson modules: identifying expository text-structures and taking notes centered on those structures. Students in the alternate treatment group were taught to make predictive inferences in narratives and to write short stories from picture prompts. Twice weekly sessions, approximately one hour in length, were administered for 4 weeks (total of 15 treatment sessions) on a university campus. Outcomes were measured through: participant identification of text-structure type in a reading passage, number of idea units related to text-structures in notes, and a multiple-choice reading comprehension task. The authors reported non-significant group differences on all outcome measures, p > 0.05.

Ukrainetz (2019) taught students to take notes and verbally rehearse using information from expository texts. This study strategy combination, *Sketch and Speak*, was designed to enhance expository comprehension and reporting skills in late elementary school-aged students with LLD. A total of 44 participants matched on performance-level were randomly assigned to either the intervention or control group and balanced to include 9 SLPs with varying levels of experience as treatment administrators. Students in the intervention group (n=22) received
treatment in 30-minute sessions twice a week for 3 weeks in one-on-one or paired treatment setting (total of 6 sessions). The majority of students received one-on-one intervention with only two sets of pairs due to SLP time constraints. Following a guided read-aloud from a trained SLP, participants created brief pictographic notes to represent essential information and then generated a complete verbal sentence about the pictograph. A second session with each topic was focused on creating bulleted notes from the pictographic notes and re-generating complete verbal sentences. All sessions ended with a full oral report to increase ownership of material after cycling through reduction and expansion through note-taking. Participants in the control condition were provided “business-as-usual” services throughout the study. Outcomes were measured through the quantity and quality of notes and holistic quality of oral reports at posttest. A different expository topic was used at testing to evaluate generalization of skills to untrained topic areas. The author reported statistically significant group differences for the quality of notes (p = 0.001) favoring the intervention group, though differences in quantity and holistic oral quality were not significant.