Abstract. This meta-analysis reviews experimental and quasi-experimental studies in which upper-elementary, intermediate, and secondary students with learning disabilities learned from graphic organizers. Following an exhaustive search for studies meeting specified design criteria, 55 standardized mean effect sizes were extracted from 16 articles involving 808 participants. Students at levels ranging from grade 4 to grade 12 used graphic organizers to learn in core-content classes (English/reading, science, social studies, mathematics). Posttests measured near and far transfer. Across several conditions, settings, and features, the use of graphic organizers was associated with increases in vocabulary knowledge, comprehension, and inferential knowledge. Mean effect sizes varied from moderate to large based on type of measure, type of graphic organizer, and subject area. Conclusions, implications for future research, and practical recommendations are presented.

students in understanding increasingly abstract concepts (Bos & Vaughn, 2002; Hughes et al., 2003; Ives & Hoy, 2003; Kim et al., 2004; Nesbit & Adesope, 2007; Rivera & Smith, 1997).

**What Are Graphic Organizers?**

GOs are visual and spatial displays that make relationships between related facts and concepts more apparent (Gajria et al., 2007; Hughes et al., 2003; Kim et al., 2004). They are intended to promote more meaningful learning and facilitate understanding and retention of new material by making abstract concepts more concrete and connecting new information with prior knowledge (Ausubel, 1968; Mayer, 1979). While there is inconsistency in the definitions of types of GOs (Rice, 1994), we were able to classify all the studies on students with LD where effects could solely be attributed to the GO into the following five general categories.

**Cognitive Mapping.** Cognitive mapping assists in making major ideas and relationships explicit by using "lines, arrows, and spatial arrangements to describe text content, structure, and key conceptual relationships" (Darch & Eaves, 1996, p. 310). By graphically displaying important information, cognitive mapping may support and develop organizational strategies essential to both reading and writing of content area material by bridging the gap between idea organization and written language structures (Pehrssen & Denner, 1988). Further, Boyle and Yeager (1997) pointed out that minimizing sentences and details in the GO is an essential component of cognitive mapping. They recommend using keywords and simple drawings rather than complex sentences or elaborate drawings. Teachers often make cognitive mapping GOs to use as advance organizers for challenging text. Students also can fill in teacher-prepared blank cognitive mapping GOs during or after attending to challenging verbal material. An example of a cognitive mapping GO is found in Figure 1.

**Semantic Mapping.** Semantic mapping (SM) is a heuristic that enables students to recognize relevant information from lecture and text (e.g., main ideas, important supporting details), delete isolated details that may not be relevant to overall understanding, and highlight key concepts that may have not been fully developed in a lecture or text (Bos & Anders, 1990).

Unlike cognitive mapping, when using SM, students and the teacher actively create a visual representation (e.g., relationship map or web) to represent the rela-
relationships among concepts. Typically, concepts are listed, and the teacher and students make predictions about how the concepts can be arranged to demonstrate those relationships on the GO (Bos & Anders, 1992).

A well-made GO consists of a superordinate concept (e.g., main idea, topic) placed in an oval in the middle or top of the page. Coordinate concepts (e.g., categories representing related concepts) are placed in ovals surrounding or underneath the superordinate concept and connected by lines. Coordinate concepts can include a variety examples, functions, or characteristics of the superordinate concept. Finally, subordinate concepts (e.g., concepts representing the coordinate concept) are listed below each coordinate concept (Bos & Anders, 1990, 1992; Pearson & Johnson, 1978). Figure 2 provides an example of an SM GO.

**Semantic Feature Analysis.** Semantic feature analysis (SFA) is similar to SM by helping students recognize relevant information from lecture and text. This is done through a presentation of related concept characteristics in a matrix form. In SFA, unrelated concepts can be inferred directly from the chart (Darch &

---

**Figure 2.** Semantic mapping example (social studies).

![Semantic Mapping Example](image)

**Figure 3.** Semantic feature analysis example (science).

<table>
<thead>
<tr>
<th></th>
<th>Comparison of Dog Breeds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basset Hound</td>
</tr>
<tr>
<td>Energetic</td>
<td>-</td>
</tr>
<tr>
<td>Weekly Grooming</td>
<td>-</td>
</tr>
<tr>
<td>Good with Kids/Other Pets</td>
<td>+</td>
</tr>
</tbody>
</table>

*Note. + = feature present; - = feature not present; ? = not sure.*
Typically, a relationship matrix is constructed with vocabulary representing the coordinate concepts placed along the top of the matrix, and the vocabulary representing the subordinate concepts placed along the side (Bos & Anders, 1990). The teacher and students can then make predictions and confirmations of relationships (e.g., related, not related, not sure) between the coordinate and subordinate concepts (Bos & Anders, 1992). The superordinate concept serves as the title. Figure 3 provides an example of an SFA GO.

Syntactic/Semantic Feature Analysis. Syntactic/semantic feature analysis (SSFA) is nearly identical to SFA but with the addition of cloze-type sentences written based on the matrix (Bos & Anders, 1990). Cloze sentences contain blank spaces replacing new vocabulary words. Students must use the context of the sentence and the SFA matrix to fill in the blanks. An example of an SSFA GO is found in Figure 4.

Visual Display. Visual displays present concepts or facts spatially, in a computationally efficient manner. That is, relationships between concepts are made apparent and clear by their location on the display. According to Hughes et al. (2003), in a visual display, facts or concepts are typically presented in one of five ways: temporal (e.g., timeline), spatial (e.g., decision tree), sequential (e.g., flowchart), hierarchal (e.g., taxonomy), or comparative (e.g., Venn diagram). An example of a visual display GO is found in Figure 5.

Previous Research

Several groups of researchers have conducted reviews and meta-analyses of the effectiveness of using GOs with nondisabled students (e.g., Ausubel, 1968; Kulhavy, Stock, & Caterino, 1994; Mayer, 1979; Moore & Readence, 1984; Nesbit & Adesope, 2007; Robinson, Katayama, Dubols, & Devaney, 1998). Based on these examinations of both the benefits of GOs and the effective design of GOs, several key findings are consistently replicated: (a) students with low verbal ability gain more from GOs than students with high verbal ability; (b) students with little or no prior knowledge in a subject gain more from GOs than students with an abundance of prior knowledge in a subject; (c) GOs are especially helpful in assisting students with far-transfer tasks, in addition to near-transfer tasks and factual recall; (d) GOs should be explicitly taught to students for maximum impact; (e) GOs should spatially group together or connect concepts so readers are more likely to perceive them as being interrelated and to draw

---

**Figure 4.** Syntactic/semantic feature analysis example (science).

<table>
<thead>
<tr>
<th></th>
<th>Comparison of Dog Breeds</th>
<th>Basset Hound</th>
<th>Old English Sheepdog</th>
<th>Brittany Spaniel</th>
<th>Border Collie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energetic</td>
<td></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Weekly Grooming</td>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Good with Kids/</td>
<td></td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>Other Pets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. + = feature present; - = feature not present; ? = not sure.*

- A(n) ____________ is the least energetic breed we have discussed.

- A(n) ____________ requires at least weekly grooming.

- A(n) ____________ and ____________ are good with other pets.
perceptual inferences about their relationships; (f) GOs should not be clustered with a lot of information; readers should easily perceive the phenomena or relations that are important; (g) GOs are effective because of their computational efficiency, minimizing stress on the working memory; and (h) GOs can be effective when used before, during, or after a lesson. While these findings are promising, the vast majority of the studies reviewed used college students as participants, and few comparisons were made to students identified as LD.

Two research syntheses of school-aged children with LD (Gajria et al., 2007; Kim et al., 2004) have focused on GOs. In Gajria et al., as part of an examination of several content enhancements, GOs were found to have large effects for comprehension of expository text for upper-elementary, intermediate, and secondary students with LD. Likewise, Kim et al. found large effects for GOs on the reading comprehension of upper-elementary, intermediate, and secondary students with LD.

Drawing definitive conclusions from these reviews is problematic for several reasons, however. First, the reviews did not take sample sizes into consideration when calculating effect sizes and comparing studies. Estimations of effect were based on individual study characteristics without standardization. The use of meta-analysis allows for statistical standardization of findings, resulting in numerical values that are interpretable in a consistent pattern across all studies, thus controlling for an individual study’s sample size (Lipsey & Wilson, 2001). Furthermore, according to Lipsey and Wilson, meaningful effects and relationships where multiple quantitative studies with varying sample sizes agree, as well as differential effects of study differences, are more likely to be identified by meta-analytic procedures than by less systematic and analytic approaches.

Second, the reviews focused solely on factual comprehension measures, not on vocabulary, inference, or relational comprehension. While information on factual comprehension is important, it may be useful to
measure the utility of GOs for students with LD on multiple constructs rather than only on the ability to answer fact-level questions about a topic. The utility of an intervention (e.g., GO) on multiple constructs is considered important for high-quality research (Gersten, Baker, & Lloyd, 2000). Thus, the effectiveness of an intervention should be examined on all possible constructs.

Third, no systematic analysis of effects was conducted by type of measure (e.g., near or far transfer), type of GO, subject area, or student stage of attending to verbal material. This lack of clarity makes it difficult to identify why the GO interventions were so effective. A more systematic analysis of GO studies may allow for more precise explanations of effect based on multiple criteria, and lead to a more robust understanding of how and why GOs are effective for students with LD.

In the present study, we conducted a meta-analysis of GO research to address the following questions:
1. What are the overall effects of GOs on the posttest performance of students with LD?
2. Do these effects maintain over time?
3. Are there differential effects by type of measure (near or far transfer)?
4. Are there differential effects by type of GO?
5. Are there differential effects by subject area?
6. Are there differential effects by stage of attending to verbal material (before, during, after instruction)?

Method

Literature Search Procedure

A three-step process was used to identify studies using GOs with upper-elementary, intermediate, and secondary students with LD. First, we conducted a comprehensive computerized search of PsycInfo, ERIC, and Social Science Citation Index databases for studies from 1975 (e.g., year LD was officially recognized by passage of P.L. 94-142) to October 2009, using a list of search terms generated from previous reviews of GO studies (e.g., Horton et al., 1993; Moore & Readence, 1984; Nesbit & Adesope, 2006). We used the following combination of descriptors: graphic organ*, expository, verbal, learning disab*, concept map, cognitive map, adolesc*, semantic map, semantic feature analysis, visual display.

Second, we conducted ancestral searches of identified articles, as well as the two most recent reviews of content enhancements used with students with LD (e.g., Gajria et al., 2007; Kim et al., 2004).

Finally, we conducted hand searches of the following journals to locate the most recent literature: Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, The Journal of Special Education, Learning Disability Quarterly, Learning Disabilities Research & Practice, Remedial and Special Education, and Reading Research Quarterly. This process yielded a total of 27 published articles to analyze, many (e.g., 20) including more than one study.

Inclusion Criteria

We used six criteria to evaluate the appropriateness of each study. First, the study must have included a dependent measure of near or far transfer of verbal (e.g., text or lecture) material and a GO as the independent variable. Studies with a mnemonic illustration rather than GO as the content enhancement (e.g., Brigham, Scruggs, & Mastropieri, 1995; Mastropieri, Scruggs, & Levin, 1987) were excluded, as effects could not be attributed solely to the GO. Likewise, studies including GOs associated with Content Enhancement Routines (Bulgren, Deshler, Schumaker, & Lenz, 2000; Bulgren, Lenz, Schumaker, Deshler, & Marquis, 2002; Bulgren, Marquis, Lenz, Schumaker, & Deshler, 2009; Bulgren, Schumaker, & Deshler, 1988; Scanlon, Deshler, & Schumaker, 1996) were excluded, because effects could not be attributed solely to the GO, even though GOs were vital components of each of these routines. Further discussion about Content Enhancement Routines is included in the Limitations section of this meta-analysis.

Second, the study must have taken place in upper-elementary-, intermediate-, or secondary-level classrooms (e.g., grades 4-12). This grade range was selected because it represents a time when curricula typically become more complex and students are required to learn primarily through didactic lecture and expository text presentation (Fletcher et al., 2007; Hughes et al., 2003; Minskoff & Allsopp, 2003).

Third, based on the recommendations of Rosenthal (1994) and Lipsey and Wilson (1993), only studies using experimental or quasi-experimental group designs with control groups were included. Therefore, single-subject research studies (e.g., Gardill & Jitendra, 1999; Idol & Croll, 1987), repeated-measures studies (e.g., Lenz, Adams, Bulgren, Pouliit, & Laraux, 2007), and single-group studies (e.g., Bergerud, Lovitt, & Horton, 1988; Boon, Fore III, Ayres, & Spencer, 2005; Horton, Lovitt, & Bergerud, 1990; Lovitt, Rudsit, Jenkins, Pious, & Benedetti, 1986; Sinatra, Stahl-Gemake, & Berg, 1984; Sturm & Rankin-Erickson, 2002) were excluded.

Fourth, the study must have provided sufficient quantitative information (e.g., group means and standard deviations; F statistic) to permit calculation of an effect size (ES). One experimental study (Boyle & Welshaar, 1997) was excluded because it provided only multivariate analysis of covariance (MANCOVA) data without group means and standard deviations. According to
Hunter and Schmidt (2004), there is no algebraically equivalent method to compute a comparable ES in such an instance.

Fifth, participants in the experimental and control groups must have included students with LD. We defined LD the same way as Kim et al. (2004) in their research review and Swanson, Hoskyn, and Lee (1999) in their meta-analysis (e.g., average intelligence and poor performance in at least one academic or related behavioral area). In each of the studies included, all participants were identified as students with LD.

Finally, the study must have been published in a peer-reviewed journal and in English. This excluded any studies in Dissertation Abstracts International and unpublished studies from researchers in the field. While this criterion ensures that only the highest quality research was included in the meta-analysis (Slavin, 1995), it also represents a potential publication bias (Lipsey & Wilson, 1993, 2001). This will be discussed further in the Limitations section.

**Study Coding**

The first author coded pertinent study features, including participant characteristics (i.e., grade level, disability classification), subject area, type of GO, stated purpose, study contrasts, dependent measures, and reported findings. A graduate research assistant double-coded this information, resulting in an interrater reliability of .97. After discussion and clarification to resolve disagreements in coding, interrater reliability reached 1.00.

**Individual Effect Size Calculation**

Using methods described by Lipsey and Wilson (2001), standardized mean difference effect size was computed using pooled standard deviation. The formula used was:

\[
ES_{\text{sm}} = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{S_p},
\]

where \(\bar{X}_{G1}\) is the mean for group 1, \(\bar{X}_{G2}\) is the mean for group 2, and \(S_p\) is the pooled standard deviation (Lipsey & Wilson, p. 48). When studies only provided an F statistic, effect size was computed using a formula recommended by Thalheimer and Cook (2002). The formula used was:

\[
d = \sqrt{F \left( \frac{n_t + n_c}{n_t n_c} \right) \left( \frac{n_t + n_c}{n_t n_c - 2} \right)},
\]

where \(n_t\) is the number of treatment subjects, and \(n_c\) is the number of control subjects.

Next, to correct for upwardly biased effect sizes due to small samples, a Hedges correction (Hedges, 1981; Lipsey & Wilson, 2001) was utilized. The unbiased effect size estimate was computed using the following formulae:

\[
ES'_{\text{sm}} = \left[ 1 - \frac{3}{4N - 9} \right] ES_{\text{sm}},
\]

\[
SE_{\text{sm}} = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1} n_{G2}} + \frac{(ES_{\text{sm}})^2}{2(n_{G1} + n_{G2})}},
\]

\[
w_{\text{sm}} = \frac{1}{SE_{\text{sm}}} = \frac{2n_{G1} n_{G2} \left(n_{G1} + n_{G2}\right)}{2(n_{G1} + n_{G2})^2 + n_{G1} n_{G2}(ES'_{\text{sm}})^2},
\]

where \(N\) is the total sample size, \(ES_{\text{sm}}\) is the biased standardized mean difference, and

\[
w_{\text{sm}} = \frac{1}{SE_{\text{sm}}}
\]

is the inverse variance weight used to calculate the weighted mean effect size. According to Hedges, Shymansky, and Woodworth (1989), the inverse variance weight is a better approach to accounting for the sample size of a given study than the simpler approach of weighting by sample size.

**Outliers**

Prior to analyzing the weighted mean effect size, extreme effect sizes that may have disproportionate influence on the analysis were eliminated (Lipsey & Wilson, 2001). Based on the recommendation of Burns (2004), eliminated outliers were effect sizes that were greater than 1.5 times the mean effect size.

**Data Analysis**

Following transformations and outlier elimination, data were analyzed by computing the weighted mean effect size:

\[
ES = \frac{\sum (w \times ES)}{\sum w}
\]

where \(\Sigma (w \times ES)\) is the summed product of the effect size and inverse variance weight, and \(\Sigma w\) is the summed inverse variance weight.

Next, the standard error of the mean effect size was computed using the following formula:

\[
se_{ES} = \sqrt{\frac{1}{\Sigma w}},
\]

where

\[
\sqrt{\frac{1}{\Sigma w}}.
\]
is the square root of one divided by the summed inverse variance weight. The z-test for the weighted mean effect size was then computed by dividing the mean effect size by the standard error of the mean effect size, or in statistical notation:

\[ Z = \frac{\bar{ES}}{SE_{ES}} \]

Finally, using the z-test, the 95% confidence interval for the weighted mean effect size was computed using the following formulae:

- \( \text{Lower} = \bar{ES} - 1.96(\text{seg}) \)
- \( \text{Upper} = \bar{ES} + 1.96(\text{seg}) \)

**Homogeneity Analysis**

Homogeneity analysis tests whether the assumption that all of the effect sizes estimate the same population mean is reasonable (Hunter & Schmidt, 2004). Lipsey and Wilson (2001) contend that single mean effect sizes by themselves are not sufficient descriptors of the distribution. Therefore, we computed a Q statistic to test homogeneity, using the formula:

\[ Q = \sum (w \times ES^2) - \left( \frac{\sum (w \times ES)}{\sum w} \right)^2 \]

The Q statistic is distributed as a chi-square with degrees of freedom (df) equaling number of ESs – 1. In our analysis of all studies, the critical value for a chi-square with df = 47 and \( p = .05 \) is 64. Because our calculated Q statistic (57.2) is less than this critical value, we can fail to reject the null hypothesis of homogeneity and assume a fixed-effects model, under which the variability across effect sizes does not exceed what would be expected based on sampling error. However, Lipsey and Wilson (2001) warned that a nonsignificant Q statistic “does not always provide great confidence that a fixed effects model is justified” (p. 117). Because we do not have a large number of effect sizes and the corresponding samples are relatively small, the Q statistic may not have sufficient statistical power. Therefore, we also fit a random-effects model, which assumes sampling error plus other sources of variability are randomly distributed (Lipsey & Wilson). The random-effects model is also a more conservative estimate than the fixed-effects model of differences between moderating variables (Hunter & Schmidt, 2004).

Whereas the fixed-effects model weights each study by the inverse of the sampling variance, the random-effects model weights each study by the inverse of the sampling variance plus a constant that represents the variability across the population effects (Lipsey & Wilson, 2001). The formulae are as follows:

\[ w_i = \frac{1}{se_i^2 + \hat{\nu}_\theta} \]

\[ \hat{\nu}_\theta = \frac{Q - k - 1}{\sum w - \left( \frac{\sum w^2}{\sum w} \right)} \]

where \( \hat{\nu}_\theta \) is the random-effects variance component. We reran the analysis using this new weight to fit the random-effects model.

The preceding analyses were conducted for posttest measures, maintenance measures, and differential effects of individual levels of independent variables (e.g., GO type, subject area) and dependent variables (e.g., near- or far-transfer measures). Because it was possible to fail to reject the null hypothesis of homogeneity, we will present results for both the fixed-effects and random-effects model for comparison. Cohen’s (1988) criteria for interpreting strength of effect sizes (small \( ES < .20 \), medium \( ES = .50 \), large \( ES > .80 \)) were used to gauge the magnitude of the findings in this analysis.

**Results**

A total of 55 unique posttest effect sizes were extracted from studies in 16 published articles meeting our inclusion criteria. For the purposes of this analysis, each unique effect size was considered an individual estimate of effect. (Included articles are marked with an asterisk in the References section.) In addition, eight of the published articles included maintenance data rendering 29 more unique effect sizes. Table 1 includes detailed information on each study, participants, variables, measures, and individual effect sizes.

**Instructional Context**

Each of the studies included instruction on the use of a GO. The majority of studies (Anders, Bos, & Filip, 1984; Bos & Anders, 1990, 1992; Boyle, 1996, 2000; Darch & Carnine, 1986; Darch, Carnine, & Kame’enui, 1986; Darch & Eaves, 1986; Darch & Gersten, 1986; DiCecco & Gleason, 2002; Englert & Mariage, 1991; Griffin, Simmons, & Kame’enui, 1991; Hudson, 1996; Ives, 2007; Reyes, Gallego, Duran, & Scanlon, 1989) incorporated aspects of direct, explicit instruction (e.g., modeling, prompted practice). The authors of the remaining study (Bos, Anders, Filip, & Jaffe, 1989) reported that written guidelines for teaching the GO were developed; however, these guidelines were not included in the article.
### Table 1
Independent Variables, Dependent Measures, and Effect Sizes of Individual Experiments

<table>
<thead>
<tr>
<th>Subject/GO Type/Study/Participants</th>
<th>Control Condition</th>
<th>Near or Far Transfer</th>
<th>Dependent Measure</th>
<th>Effect Size Posttest M</th>
<th>Maintenance M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SOCIAL STUDIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Semantic Mapping</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 5: 26 upper-elementary students with LD</td>
<td>Normative</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.42</td>
<td>.62*</td>
</tr>
<tr>
<td></td>
<td>Normative</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.30</td>
<td>.06</td>
</tr>
<tr>
<td><em>Semantic Feature Analysis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anders, Bos, &amp; Filip (1984); 62 high school students with LD</td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.71***</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.49***</td>
<td>N/A</td>
</tr>
<tr>
<td>Bos, Anders, Filip, &amp; Jaffe (1989); 50 high school students with LD</td>
<td>Dictionary method</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.67***</td>
<td>.00</td>
</tr>
<tr>
<td><strong>SM, SFA, SSFA Combination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 1: 42 upper-elementary students with LD</td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.81</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>.50</td>
<td>1.15</td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 3: 47 upper-elementary students with LD</td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.46</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.28</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Visual Display</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darch &amp; Carnine (1986); 24 upper-elementary students with LD</td>
<td>Text-only</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.73***</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Text-only</td>
<td>Far</td>
<td>Researcher-generated far-transfer comprehension Test</td>
<td>.64</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*continued next page*
Table 1 continued
Independent Variables, Dependent Measures, and Effect Sizes of Individual Experiments

<table>
<thead>
<tr>
<th>Subject/GO Type/Study/ Participants</th>
<th>Control Condition</th>
<th>Near or Far Condition</th>
<th>Dependent Measure</th>
<th>Effect Size Posttest M</th>
<th>Maintenance M</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOCIAL STUDIES continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Visual Display continued</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darch, Carnine, &amp; Kame'enui (1986); 84 junior high school students with LD</td>
<td>Directed reading</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.79***</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Directed reading</td>
<td>Far</td>
<td>Researcher-generated far-transfer comprehension test</td>
<td>.82***</td>
<td>N/A</td>
</tr>
<tr>
<td>DiCecco &amp; Gleason (2002); 24 junior high students with LD</td>
<td>Guided discussion</td>
<td>Near</td>
<td>Researcher-generated fact quizzes</td>
<td>.05</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Guided discussion</td>
<td>Far</td>
<td>Relational statements in two written essays</td>
<td>.94***</td>
<td>N/A</td>
</tr>
<tr>
<td>Hudson (1996); 21 junior high students with LD</td>
<td>Note-taking guide</td>
<td>Near</td>
<td>Researcher-generated knowledge test</td>
<td>1.58***</td>
<td>2.37***</td>
</tr>
<tr>
<td></td>
<td>Note-taking guide</td>
<td>Far</td>
<td>Researcher-generated inference test</td>
<td>1.00**</td>
<td>2.08***</td>
</tr>
<tr>
<td>SCIENCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semantic Mapping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1990); 61 junior high students with LD</td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.27***</td>
<td>.94**</td>
</tr>
<tr>
<td></td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.33***</td>
<td>.54</td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 6: 22 junior high students with LD</td>
<td>Normative</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>2.16***</td>
<td>.78*</td>
</tr>
<tr>
<td>Reyes, Gallego, Duran, &amp; Scanlon (1989); 61 junior high students with LD</td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.27***</td>
<td>.94**</td>
</tr>
<tr>
<td></td>
<td>Definition instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.33***</td>
<td>.47</td>
</tr>
</tbody>
</table>

continued next page
### Table 1 continued

**Independent Variables, Dependent Measures, and Effect Sizes of Individual Experiments**

<table>
<thead>
<tr>
<th>Subject/GO Type/Study/ Participants</th>
<th>Control Condition</th>
<th>Near or Far Condition</th>
<th>Dependent Measure</th>
<th>Effect Size Posttest M</th>
<th>Effect Size Maintenance M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCIENCE continued</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Semantic Feature Analysis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1990); 61 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.03**</td>
<td>.66*</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.47***</td>
<td>.44</td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 6: 22 junior high students with LD</td>
<td>Normative</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.17</td>
<td>.73*</td>
</tr>
<tr>
<td>Reyes, Gallego, Duran, &amp; Scanlon (1989); 61 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.03***</td>
<td>.66*</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.47***</td>
<td>.44</td>
</tr>
<tr>
<td><em>Syntactic/Semantic Feature Analysis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1990); 61 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>.64</td>
<td>1.38***</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.18**</td>
<td>1.40***</td>
</tr>
<tr>
<td>Reyes, Gallego, Duran, &amp; Scanlon (1989); 61 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>.64</td>
<td>1.38***</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.18***</td>
<td>1.40***</td>
</tr>
<tr>
<td><em>SM/SFA/SSFA Combination</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 2: 61 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.22</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>.92</td>
<td>1.01</td>
</tr>
<tr>
<td>Bos &amp; Anders (1992); Study 4: 53 junior high students with LD</td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.51</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>Definition; instruction</td>
<td>Near</td>
<td>Researcher-generated vocabulary test</td>
<td>.83</td>
<td>.79</td>
</tr>
</tbody>
</table>

*continued next page*
Table 1 continued
*Independent Variables, Dependent Measures, and Effect Sizes of Individual Experiments*

<table>
<thead>
<tr>
<th>Subject/GO Type/Study/ Participants</th>
<th>Near or Far Condition/ Transfer</th>
<th>Dependent Measure</th>
<th>Effect Size Posttest M</th>
<th>Maintenance M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCIENCE continued</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Visual Display</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darch &amp; Eaves (1986); 22 high school students with LD</td>
<td>Text-only Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.29***</td>
<td>.35</td>
</tr>
<tr>
<td>Text-only Near</td>
<td>Researcher-generated far-transfer comprehension test</td>
<td>.64</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Griffin, Simmons, &amp; Kame'enui (1991); 28 upper-elementary students with LD</td>
<td>No visual display Near</td>
<td>Researcher-generated comprehension test</td>
<td>.51</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>SCIENCE/SOCIAL STUDIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Visual Display</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darch &amp; Gersten (1986); 24 high school students with LD</td>
<td>Basal instruction Near</td>
<td>Researcher-generated comprehension test</td>
<td>1.72***</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Semantic Feature Analysis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Darch &amp; Gersten (1986); 24 high school students with LD</td>
<td>Dictionary method Near</td>
<td>Researcher-generated vocabulary test</td>
<td>1.66***</td>
<td>.00</td>
</tr>
<tr>
<td><strong>ENGLISH/LANGUAGE ARTS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cognitive Mapping</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boyle (1996); 30 junior high school students with LD or MMR</td>
<td>Typical reading instruction Near</td>
<td>Formal reading inventory</td>
<td>.34</td>
<td>N/A</td>
</tr>
<tr>
<td>Typical reading instruction Near</td>
<td>Below-grade-level literal comprehension test</td>
<td>.87**</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Typical reading instruction Near</td>
<td>On-grade-level literal comprehension test</td>
<td>1.33***</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Typical reading instruction Far</td>
<td>Below-grade-level inferential comprehension test</td>
<td>.76**</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Typical reading instruction Far</td>
<td>On-grade-level inferential comprehension test</td>
<td>.96***</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

continued next page
Table 1 continued

**Independent Variables, Dependent Measures, and Effect Sizes of Individual Experiments**

<table>
<thead>
<tr>
<th>Subject/GO Type/Study/ Participants</th>
<th>Near or Far Condition</th>
<th>Dependent Measure</th>
<th>Effect Size (Posttest M)</th>
<th>Maintenance M</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGLISH/LANGUAGE ARTS continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cognitive Mapping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastropieri &amp; Peters (2003); 20 junior high students with LD</td>
<td>List illustration Near</td>
<td>Research-generated featured item recall test</td>
<td>1.42***</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>List illustration Far</td>
<td>Research-generated featured item recall test</td>
<td>1.06**</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Semantic Mapping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Englert &amp; Mariage (1991); 28 upper-elementary students with LD</td>
<td>Typical reading instruction Near</td>
<td>Written free recall</td>
<td>1.84***</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Visual Display</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boyle (2000); 24 high school students with LD or MMR</td>
<td>No training Near</td>
<td>Literal comprehension test</td>
<td>1.14***</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>No training Near</td>
<td>Relational comprehension test</td>
<td>.91**</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>No training Far</td>
<td>Inferential comprehension test</td>
<td>.51</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>MATHEMATICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Visual Display</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ives (2007); Study 1: 30 high school students with LD</td>
<td>Control Near</td>
<td>Teacher-generated test</td>
<td>.67*</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Control Near</td>
<td>Researcher-generated test (concepts)</td>
<td>1.06**</td>
<td>.94**</td>
</tr>
<tr>
<td></td>
<td>Control Far</td>
<td>Researcher-generated test (system solving)</td>
<td>.16</td>
<td>.00</td>
</tr>
<tr>
<td>Ives (2007); Study 2: 20 high school students with LD</td>
<td>Control Near</td>
<td>Researcher-generated test</td>
<td>.49</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**p < .001; **p < .05; *p < .1; all GOs were created by the researchers.**
Generally, instruction for the experimental groups included one to two sessions focused solely on how to use the GO, one to two sessions of prompted practice using the GO, and independent student use of the GO for the remainder of sessions. During the initial sessions, the teacher or researcher presented the GO to students and described how it illustrated relationships. For example, Darch and Carnine (1986) presented their visual display via an overhead projector, and students followed along while the teacher used a script to describe the various cells in the display and the interrelationships between them.

The following sessions generally included the instructor explicitly guiding the students in creating or filling out the GO. For example, Bos and Anders (1990) explicitly prompted the students in each step of creating a hierarchical semantic map from a vocabulary list. This level of assistance was gradually faded. For instance, Darch and Gersten (1986) first presented a visual display with all the cells labeled, and prompted the entire group in answering questions about specific facts in the GO. The researchers followed this by guiding the students through a visual display that did not provide cell labels. Finally, individual students were prompted in labeling blank visual displays. Instruction in the remaining sessions generally focused on independent use of the GO by the students, in addition to text or lecture presentations. However, in each of the visual display studies all of the content was presented solely through the GO.

Each of the interventions lasted between one and seven weeks, with an additional one to four weeks between posttest and maintenance measures. All of the studies were conducted in a resource classroom during or after the regular school day.

What Are the Overall Effects of GOs on the Posttest Performance of Students With LD?

After the removal of six outliers, there was a large overall standardized effect of GOs on the posttest performance (e.g., multiple-choice comprehension, vocabulary, written recall) of students with LD across all studies (ES = .91, SE = .062) for both random- and fixed-effects models and a 95% confidence interval of .79, 1.03 for the random effects model. Table 2 provides the full comparison between the random- and fixed-effects models.

Do These Effects Maintain Over Time?

Twenty-nine studies included maintenance measures. In each of the studies, measures consisted of multiple-choice comprehension or vocabulary items. These measures were given to students one to four weeks after the conclusion of the intervention.

The test of homogeneity for overall maintenance effects produced a nonsignificant Q statistic (Q = 24.49, cv = 35.17). Therefore, similar to the overall posttest effects, results of both the random- and fixed-effects models are reported. After removal of five outliers, there was a medium overall effect for maintenance across all studies (ES = .56, SE = .074) with a 95% confidence interval of .41, .70 for the random-effects model. Table 3 provides the full comparison between the random- and fixed-effects models.

Differential Effects

Based on number of effect sizes and significant Q-values, the remaining analyses of differential effects are

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Overall ES for Fixed and Random Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*p < .001.
reported as the random-effects model. This provides a more conservative estimate of effect.

Are There Differential Effects by Type of Measure (Near or Far Transfer)? Forty-five individual estimates of effect were calculated for near-transfer measures at posttest and 27 individual estimates of effect for maintenance. In all but two articles, near-transfer measures consisted of researcher-generated multiple-choice questions on material directly covered in the lessons. In the remaining two articles, Boyle (1996) used a standardized measure of reading comprehension and Englert and Mariage (1991) used a measure of written free recall, respectively. Overall mean effect size was 1.07 for posttest and .78 for maintenance.

Ten individual estimates of effect were calculated for far-transfer measures at posttest and two individual estimates of effect for maintenance. These measures tested students’ ability to apply knowledge to situations not directly covered in the text or lecture. For example, Hudson (1996) included the question “Name one way the environment influenced the culture of the Arctic tribes” (p. 81). The teacher never stated the causal relation between environment and culture; only facts about environment and culture were stated.

### Table 3

<table>
<thead>
<tr>
<th>Maintenance ES for Fixed and Random Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
</tr>
<tr>
<td>Fixed Model</td>
</tr>
<tr>
<td>Random Model</td>
</tr>
</tbody>
</table>

*p < .001.

### Table 4

Near- and Far-Transfer – Random-Effects Model

<table>
<thead>
<tr>
<th></th>
<th>Posttest</th>
<th>Maintenance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ES</strong></td>
<td><strong>95% CI</strong></td>
<td><strong>ES</strong></td>
</tr>
<tr>
<td><strong>Near Transfer</strong></td>
<td>1.065</td>
<td>.94</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td><em>n = 45</em></td>
<td>(n = 45)</td>
<td></td>
</tr>
<tr>
<td><strong>Far Transfer</strong></td>
<td>.6127</td>
<td>.36</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td><em>n = 10</em></td>
<td>(n = 10)</td>
<td></td>
</tr>
</tbody>
</table>

Note. *n* = number of ESs.
As for near-transfer, far-transfer measures consisted of researcher-generated multiple-choice questions in all but one study. Boyle (1996) used a standardized measure for far transfer. The overall mean effect size was .61 for posttest and .69 for maintenance. Table 4 provides the full comparison between near- and far-transfer measures.

Are There Differential Effects by Type of GO? The types of GOs used in the studies matched the definitions in the introduction to this analysis (e.g., cognitive mapping, SM, SFA, SSFA, visual display). However, in one article containing eight studies (Bos & Anders, 1992), the researchers used a combination of SM, SFA, and SSFA. The method they utilized to present their results prohibited disaggregation of the findings. Therefore, a sixth category (SM/SFA/SSFA Combination) was added to the analysis. Large posttest effects (e.g., .74-1.2) were found for all types of GOs except visual displays. Visual displays had a moderate effect (e.g., .74). There were no statistically significant differences between GOs with large posttest effects. For maintenance measures, SSFA and SM/SFA/SSFA Combination had significantly larger effects than the other GO types (e.g., 1.39, 1.01). Table 5 provides the full comparison between types of GOs.

Are There Differential Effects by Subject Area? Posttest effects were calculated for the subject areas of English/writing/reading, mathematics, science, and social studies. Large posttest effects were found for all subject (e.g., .96-1.05) areas except mathematics (e.g., .59). Mathematics had a moderate posttest effect that was significantly smaller than the other subject areas. Maintenance effects were calculated for mathematics, science, and social studies. Science had a large maintenance effect (e.g., .80) that was significantly larger than the moderate effects for mathematics and social studies. Table 6 provides the full comparison between GOs by subject area.

Are There Differential Effects by Stage of Attending to Verbal Material (Before, During, After Instruction)? There was not enough information to quantify differential effects by stage of instruction. All
but one study included GOs before and during instruction, and not enough information was provided to disaggregate these data. One study, Englert and Mariage (1991), used GOs after instruction. The unstandardized effect size for a near-transfer, written free recall measure was large \( (ES = 1.84) \).

**Discussion**

As was the case in previous research syntheses (e.g., Gajria et al., 2007; Kim et al., 2004; Moore & Readence, 1984), findings from this meta-analysis indicate that GOs improve the factual comprehension of upper-elementary, intermediate, and secondary students with LD. Unlike these previous reviews, this analysis also indicates that GOs may improve vocabulary and inference/relational comprehension for students with LD.

What Are the Overall Effects of GOs on the Posttest Performance of Students With LD?

Overall, there was a large mean effect for posttest performance \( (ES = .91, SE = .06) \) of students with LD using both the fixed-effects and random-effects models. The immediate posttest performance spanned multiple constructs, including multiple-choice factual comprehension, vocabulary, and written recall requiring relational comprehension. This suggests that GOs are effective in not only improving basic skills (e.g., factual recall) but also higher-level skills (e.g., inference). This finding is consistent with the theories of Ausubel (1968) and Mayer (1979) that GOs may especially assist lower ability learners in both basic and higher level skills by creating an easier context for assimilating information into their memory.

Do These Effects Maintain Over Time?

There was a moderate mean effect for maintenance \( (ES = .56, SE = .07) \) of students with LD. The significant drop-off from posttest to maintenance is consistent with the findings of the other GO research syntheses. The drop has been attributed to lack of clarity regarding the duration and length of intervention sessions needed to positively affect maintenance (Gajria et al., 2007; Gersten, Fuchs, Williams, & Baker, 2001). The relatively short duration of the intervention studies (e.g., 1-7 weeks) may not have provided sufficient instruction time for students to use GOs independently. However, a closer look at effects by type of GO shows that effect sizes for visual displays and SSFA were larger for maintenance than posttest. This may lend support to the visual argument hypothesis (Waller, 1981), which posits that GOs that are structured in a way that facilitates understanding and perception of concept relationships are superior to more complicated GOs that may require instruction to recognize conceptual relationships (Dexter, 2010). The visual displays and SSFA were more computationally efficient than the other GOs. That is,

---

**Table 6**

*Subject Area – Random-Effects Model*

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Posttest</th>
<th></th>
<th></th>
<th>Maintenance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ES</td>
<td>Lower</td>
<td>Upper</td>
<td>ES</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>English/Writing/Math</td>
<td>.9612</td>
<td>.72</td>
<td>1.20</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>n = 11</td>
<td></td>
<td>n = 11</td>
<td></td>
<td></td>
<td>n = 0</td>
<td>N/A</td>
</tr>
<tr>
<td>Mathematics</td>
<td>.5942</td>
<td>.21</td>
<td>.98</td>
<td>.4559</td>
<td>.07</td>
<td>.99</td>
</tr>
<tr>
<td>n = 4</td>
<td></td>
<td>n = 4</td>
<td></td>
<td></td>
<td>n = 2</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>1.052</td>
<td>.88</td>
<td>1.23</td>
<td>.8035</td>
<td>.64</td>
<td>.97</td>
</tr>
<tr>
<td>n = 23</td>
<td></td>
<td>n = 23</td>
<td></td>
<td></td>
<td>n = 20</td>
<td></td>
</tr>
<tr>
<td>Social Studies</td>
<td>1.037</td>
<td>.85</td>
<td>1.22</td>
<td>.6535</td>
<td>.38</td>
<td>1.03</td>
</tr>
<tr>
<td>n = 19</td>
<td></td>
<td>n = 19</td>
<td></td>
<td></td>
<td>n = 8</td>
<td></td>
</tr>
</tbody>
</table>

*Note. n = number of ESs.*
they were simple enough for students to recognize conceptual relationships without teacher instruction. This may explain why maintenance effects were larger for these types of GOs.

Are There Differential Effects by Type of Measure (Near or Far Transfer)?

This meta-analysis also separated results into near-transfer and far-transfer measures. Near-transfer results (i.e., measures applying knowledge to situations directly covered in the text or lecture) indicate that GOs are effective strategies for improving factual recall, factual and relational comprehension, and vocabulary knowledge. Across all near-transfer studies, the mean effect size was large ($ES = 1.07$), and maintenance effects were moderate ($ES = .78$). Students using GOs significantly outperformed their peers receiving typical classroom instruction on near-transfer measures. Interestingly, more complicated GOs requiring intensive teacher instruction (e.g., SM, SFA) resulted in the largest effects for near-transfer posttest measures. This indicates that while these GOs are difficult to understand independently, with appropriate instruction they are superior to less complicated GOs for immediate factual recall.

Far-transfer results (i.e., measures applying knowledge to situations not directly covered in the text or lecture) indicate that GOs may also improve the inference skills and relational knowledge of secondary students with LD. Across all far-transfer studies, the mean effect size was moderate ($ES = .61$), and maintenance effects were moderate ($ES = .69$).

It is interesting to note that for far-transfer measures, maintenance effect sizes were larger than posttest effect sizes. Previous research has indicated students with LD typically perform poorly on far-transfer tasks due to their inability to detect underlying concepts in verbal information due to difficulty assimilating verbal information with previous knowledge (Suritsky & Hughes, 1991). The results of this analysis demonstrate that GOs may bridge the gap between verbal information and prior knowledge and assist students with LD in far-transfer tasks. This finding supports Mayer's (1979) assimilation theory, which posits that GOs that assimilate material to a broader set of past experiences allow superior transfer to new situations.

The finding is also consistent with the research of Robinson and colleagues (Robinson et al., 1998; Robinson & Kiewra, 1995; Robinson & Schraw, 1994; Robinson & Skinner, 1996), comparing visual displays (e.g., tree diagrams, matrices, network charts) with traditional, non-graphic outlines. In each of the studies, groups of nondisabled college students using GOs and traditional outlines equally outperformed text-only groups in factual recall, but the GO groups significantly outperformed the outline and text-only groups in identifying concept relations and making far-transfer concept comparisons.

Are There Differential Effects by Subject Area?

This analysis also examined the effects of GOs by subject area. All subject areas had moderate to large effects for posttest and maintenance measures. The largest effects were in science ($ES = 1.05$ for posttest, $ES = .80$ for maintenance) and the smallest effects in mathematics ($ES = .59$ for posttest, $ES = .46$ for maintenance). The large effects in science may be explained by the unfamiliar, technical vocabulary and content often based on relationships between concepts (Lovitt et al., 1986). This type of content lends itself to computationally efficient GOs that make relationships explicit and clear. Also, students may rely more heavily on content enhancements like GOs when content seems strange or foreign.

The small effects for mathematics may be explained by the fact that the information was much more abstract than the other subject areas. It may also be explained by the fact that the mathematics study used a visual display, which only had a moderate overall effect. Use of GOs with mathematics concepts and solving systems of linear equations are only beginning in the field. Ives (2007) offered several implications for future research based on his initial study in this field. It will take time and more study to fully understand the effects of GOs on mathematics understanding.

Are There Differential Effects by Stage of Attending to Verbal Material (Before, During, After Instruction)?

Finally, a previous GO research synthesis (Moore & Readence, 1984) reported that GOs presented as text summaries after instruction were more effective than GOs presented before or during instruction. This finding cannot be corroborated by the present analysis because only one study (Englert & Marlage, 1991) used GOs after instruction. While this study had an extremely large effect size (1.84), more studies are needed to confirm this finding. The current analysis points to effective instruction and choice of GO to be more important than stage of attending to verbal material in effectiveness of the intervention.

Methodological Limitations

There are two methodological limitations to the conclusions of this analysis. First, there is the possibility of a publication bias. According to Smith (1980), as well as Lipsey and Wilson (1993), published articles have a larger mean effect size than unpublished studies. This bias, also known as a "file-drawer" effect, happens because studies with null findings are less likely to be
published by major journals, often leaving offending
data in a researcher's file drawer (Lipsey & Wilson,
2001). In this analysis, based on the advice of Slavin
(1995), we purposefully selected only published studies
to ensure the highest quality of research designs. While
this eliminated our ability to compare mean effect sizes
of published studies versus unpublished studies, we
were able to utilize Rosenthal's (1979) fail-safe N sta-
tistic, which was later adapted by Orwin (1983). This sta-
tistic determines “the number of studies with an effect
size of zero needed to reduce the mean effect size to a
specified or criterion level” (Lipsey & Wilson, 2001,
p. 166).

The formula is as follows:

\[ k_0 = k \left[ \frac{ES_k}{ES_c} - 1 \right], \]

where \( k_0 \) is the number of effect sizes of zero needed to
reduce the mean effect size to \( ES_c \) (criterion effect level),
\( k \) is the number of current studies, and \( ES_k \) is the cur-
rent weighted mean effect size. Using this statistic,
reduction of our overall weighted effect size of .91 to
.50 would take 45 additional unpublished studies with
an effect size of zero. While this may be a possibility,
It is unlikely that 45 additional null studies exist in
researchers' file drawers.

Second, our 55 unique effect sizes or studies were
culled from only 16 published articles. While this is
an acceptable practice (Lipsey & Wilson, 2001), it does
limit the generalizability of the findings because there
were only 21 distinct samples of students with LD
(total \( N = 808 \)). Therefore, caution should be exercised
in generalizing these findings to all intermediate and
secondary students with LD.

Individual Study Limitations

Three limitations to the individual studies warrant
consideration. First, while all of the effect sizes in the
present analysis were based on differences between a
treatment group and a control group, it is not clear if
the control conditions constituted an adequate stan-
dard to measure the effects of GO interventions
(Gersten, Baker, & Lloyd, 2000). The control conditions
in the studies used primarily typical classroom practices
(e.g., dictionary instruction) rather than more closely
comparable practices (e.g., outlines, structured over-
views). While this provides much evidence for GO
effects over typical classroom practice, it does not yield
information comparing GOs with other researched
practices (Kim et al., 2004)

Second, while the results indicate large effects for
vocabulary, inference, and comprehension, it is impor-
tant to note that all but one study used measures that
were researcher-created and closely tied to the content.
While these measures should have good content valid-
ity, there is no way to measure broader construct valid-
ity. Only Boyle (1996) used a standardized measure for
reading comprehension. This fact may limit the gen-
eralizability of these findings and questions the actual
level of understanding obtained by students in the GO
conditions.

Finally, studies using Content Enhancement
Routines were not included in this meta-analysis
because their effects could not be attributed solely to
the GO, even though GOs are a vital component of
each routine. Research on Content Enhancement
Routines has been conducted for over two decades,
focuses primarily on assisting all students, including
those with LD, thrive in the often rigorous intermedi-
ate and secondary general education content class-
rooms (Bulgren, 2006; Bulgren, Deshler, & Lenz, 2007
These routines have been shown to improve both basic
skills (e.g., factual knowledge, comprehension) and
higher-order skills (e.g., manipulation, extension, gen-
eralization) for students with LD (e.g., Bulgren et al.,
2000; Bulgren et al., 2002; Bulgren et al., 2009; Scanlon
et al., 1996). While results of these studies did not fit
into our analysis, they do serve as examples of effective
interventions utilizing GOs as a component of a larger
routine.

Implications for Practice

The major implication of this study for applied prac-
tice is consistent with assimilation theory and the
visual argument hypothesis; that is, more instruction-
intensive types of GOs (e.g., SM, SFA) are better for
immediate factual recall while more computationally
efficient GOs (e.g., visual display, SSFA) are better for
maintenance and transfer. This knowledge can help
teachers in designing GOs for initial instruction and for
re-teaching, studying, and retention purposes. For
instance, a semantic map for initial instruction, fol-
lowed by a simpler visual display for review and study
will potentially maximize the effects of recall, mainte-
nance, and far-transfer for students with LD.

Another implication for practice is that, regardless of
GO type, a teacher must explicitly teach students how
to use a given GO. Students with LD need explicit
instruction to understand how concepts are related, to
recognize differences between main and subordinate
ideas, and to put all the pieces together to make a clear
picture of the content being learned no matter how
implicit a GO may seem. A teacher's use of effective
instruction practices (e.g., modeling, corrective feed-
back, etc.) will positively impact the intervention's
effectiveness.
Conclusions and Implications for Future Research

This meta-analysis found that, in comparison to activities such as reading text passages, attending to lectures, and participating in typical classroom practice (e.g., dictionary instruction), GOs are more effective on posttest, maintenance, and transfer measures. However, this conclusion must be tempered for several reasons.

First, each of the studies took place in self-contained resource classrooms. This may not be typical for today’s upper-elementary, intermediate, and secondary students with LD. As many students with LD are now fully included with nondisabled peers in core content classes, it is important to closely examine how GOs will work in this setting. The feasibility and practicality of GOs must be examined in general education settings and recommendations for effective use put forth.

Second, there is great need for GO replication studies. Only three articles in the present meta-analysis were published in the past 10 years. More current group design, randomized control trials, is needed to fully validate the benefits of GOs across all secondary students with LD.

Finally, for student independent practice, it was not always clear from the studies if the GO was used correctly or at all. For instance, when students independently filled in a blank GO, there was no reported procedure for ascertaining if they were properly labeling main and subordinate details. Likewise, several of the studies (e.g., Anders, Bos, & Filip, 1984; Bos & Anders, 1990; Bos & Anders, 1992) reported students had a GO and text to study for the posttests. They did not include a procedure for making sure the students were actually using the GO to study. These students may have been using the text as their study guide. This lack of control may somewhat negate the attribution of effects to the GO. Future research must tightly control for these potential problem areas.

Taking the above issues into account, the evidence in this analysis should still persuade educational practitioners to make well-planned and well-instructed use of GOs. There were no significant negative effects across any of the categories of analysis and no other identified detrimental effect. A thoughtful combination of types of GOs will help make the learning process more efficient for upper-elementary, intermediate, and secondary students with LD.

References


Bulgren, J., Deshler, D. D., & Lenz, B. K. (2007). Engaging adolescents with LD in higher order thinking about history concepts...


* = Studies included in meta-analysis.

Please address correspondence about this article to: Douglas D.
Dexter, 230 CEDAR, The Pennsylvania State University, University Park, PA 16802; email:Ddd176@psu.edu