Journal of Educational Psychology
2002, Vol. 94, No. 2, 356–371

0022-0663/02/$5.00 DOI: 10.1037//0022-0663.94.2.356

The Use and Effectiveness of a Comparison Routine in Diverse Secondary Content Classrooms

Janis A. Bulgren, B. Keith Lenz, Jean B. Schumaker, Donald D. Deshler, and Janet G. Marquis
University of Kansas

The purpose of this investigation was to explore teachers' use and effectiveness of a comparison routine to help students understand comparisons of important content information in secondary content classes containing students of diverse ability. Various research methodologies, including single-subject and large-group experimental designs, were used. Measures included students' knowledge of information involving comparisons, the numbers and types of comparisons teachers used, teachers' use of the instructional routine, and teachers' and students' satisfaction with the instruction. Use of the routine led to significantly better retention and expression of information by students in the experimental condition compared with students participating in a traditional lecture-discussion format. Teachers easily learned the routine, delivered more complete instruction, and used the routine to teach regularly scheduled curricula. Teachers indicated that they were satisfied with the routine; in general, students were neither satisfied nor dissatisfied with the routine.

Academic failure by some adolescents in secondary schools is well documented (e.g., Deshler & Schumaker, l988; McDonnell, McLaughlin, & Morrison, 1997). Teachers at all levels are concerned about students who struggle with the academic demands they are expected to meet. However, they are often perplexed about how to help these students learn, especially when these students are enrolled in classes containing students with diverse abilities (Joint Committee on Teacher Planning for Students With Disabilities, 1995; Nolet & McLaughlin, 2000).

To complicate matters, academic demands at the secondary level have been intensified by the standards-based reform movement (McDonnell et al., 1997). Many content standards being adopted by states emphasize higher order thinking skills such as making comparisons, in addition to the acquisition of factual information. For example, the National Council for the Social Studies (1990) has emphasized the ability to "compare things, ideas, events, and situations on the basis of similarities and differences" (p. 10). The National Council of Teachers of Mathematics (2000) has emphasized the ability to "compare and contrast situations that are modeled by functions from various classes" (p. 296). Additionally, commentators also have emphasized the importance of using comparison structures in content classes (e.g., Bulgren & Lenz, 1996; Dickson, Simmons, & Kameenui, 1995).

This emphasis on higher order thinking is, of course, not new. Teachers have long relied on taxonomies (e.g., Bloom, 1956) to guide educational objectives such as those related to the acquisition of the skills involved in making comparisons. One area within which such skills might be taught relates to the comparison of concepts. A concept is a "category or class into which events, ideas, or objects can be grouped" (Bulgren, Deshler, & Schumaker, 1993, p. 7).

Some research has been conducted on the development and validation of instructional routines (i.e., packages of teaching practices) for teaching concepts at the secondary level in diverse classes of students. This research has been based on the Content Enhancement Approach to instruction, a new approach to secondary subject-area education (Bulgren & Lenz, 1996; Schumaker, Deshler, & McKnight, 1991) to enhance students' understanding, retention, and recall of information. For example, Bulgren, Schumaker, and Deshler (1988) developed a routine for teaching concepts, the Concept Mastery Routine (Bulgren et al., 1993) and tested its effectiveness in secondary social studies and science classes in which diverse groups of students were enrolled. The routine is based on the identification and graphic display of characteristics and examples of a concept and how they can be used to create a definition of the concept. The researchers found that students, including those with disabilities, taught with the routine had scores that were significantly higher on both specially designed tests about the concepts taught and on regularly scheduled publisher-made unit tests than when the routine was not used.

Later, the same researchers tested another routine, the Concept Anchoring Routine, designed to help students learn new, difficult

Janis A. Bulgren, B. Keith Lenz, Jean B. Schumaker, and Donald D. Deshler, Center for Research on Learning, University of Kansas. Janet G. Marquis, Institute for Life Span Studies, University of Kansas.

Preparation of this article was supported in part by National Institute of Disability and Rehabilitation Research Grant H133G90156. We acknowledge the administrators and teachers of Lawrence High School and Central Junior High School in Lawrence School District 497 in Kansas, including Ed Judd, Andrea Parson, Ruthie Rapp, and Kathy vonEnde, and the administrators and teachers of Shawnee Mission School District 512 in Kansas, including Diane Maddox, Jenny Sands, Karen Noland, Bob James, Jerry Cormack, Glenda Johnson, Shelly King, Kay Neill, Dave Taylor, and Kathy Taylor, for their help on this project.

Correspondence concerning this article should be addressed to Janis A. Bulgren, Center for Research on Learning, University of Kansas, J. R. Pearson Hall, Room 517, 1122 West Campus Road, Lawrence, Kansas 66045-3101. E-mail: jbulgren@ku.edu

concepts through analogies. The researchers found statistically significant differences in favor of the use of the routine over a traditional lecture-discussion format in a laboratory-type classroom arrangement and also in an actual secondary class taught by a subject-area teacher under typical school conditions (Bulgren, Deshler, Schumaker, & Lenz, 2000).

Such graphic devices have been popularly touted in the literature and often used by teachers because of their potential with regard to representing complex ideas and relationships in a logical, clear, and concrete fashion (e.g., Beissner, Jonassen, & Grabowski, 1993; Clarke, 1991; Horton & Lovitt, 1989; Irvin, 1990). Such clear presentations may be particularly important for students who struggle with learning, such as those with a learning disability (LD), for whom retention processes are more influenced by interference than those of normal children (Mayer, 1987) or who may have monitoring difficulties as well as poor automatization of skills related to the cognitive components involved in information processing (Swanson, 1990).

Until the Bulgren et al. (1988) study, the research on the use of graphic organizers had been relatively inconclusive with regard to their effectiveness in enhancing student learning of complex information delivered during live instruction in secondary classes. In general, reviewers of these studies concluded that the use of these graphic organizers in instruction did not produce strikingly large differences between groups (Moore & Readence, 1984). However, the best effects seemed to be achieved when students were trained and participated in constructing the graphic organizers, when graphic organizers were used with more capable students, and when they were used in conjunction with descriptive texts (Dunston, 1992).

Several studies have been conducted that included experimental designs and that focused on secondary students receiving instruction through the use of graphic organizers to help them comprehend text. Specifically, eight senior high students with LD participated in a study in which they earned significantly higher test scores when they worked with graphic organizers and their texts than when they received lectures and took notes based on the text (Doyle, 1999). In a group of three studies conducted with heterogeneous groups of students, including some students with disabilities, teacher-directed and student-directed use of an organizer in relation to reading assignments produced significantly higher test scores than self-study (Horton, Lovitt, & Bergerud, 1990). However, Bean, Singer, Sorter, and Frazee (1986) concluded that graphic organizer instruction alone was no more effective than outlining instruction for 10th-grade honors world history students.

Only a couple of studies have focused on the use of graphic organizers to make comparisons. Again, however, the graphic organizers, which comprised matrices called *relationship charts*, were used to help students comprehend written text by focusing on similarities and differences of word meanings. Using an instructional routine called *semantic feature analysis*, high school teachers and students with disabilities in resource English and social studies classes filled in relationship charts (Bos, Anders, Filip, & Jaffe, 1989). The students earned higher scores on a comprehension test taken immediately after instruction and again 6 months after the semantic feature analysis was used than when they wrote definitions of the words. In another study, Bos and Anders (1990) reported that middle school students with LD in a resource room who made comparison charts with mapping graphics earned significantly higher scores on a test taken immediately after the instruction than did students who participated in semantic feature analysis.

To summarize, although some studies have been conducted on the use of graphic organizers to help students comprehend text at the secondary level and although some impressive effects have been achieved in a few of those studies, none of these studies has focused solely on the live instruction of comparative information in general education classes at the secondary level. Furthermore, although some studies have been conducted on the live instruction of conceptual information about a single concept (Bulgren et al., 1998, 2000) and on some relatively simple comparisons (Bos & Anders, 1990; Bos et al., 1985), no studies have focused on the comparison of two complex concepts in which the major characteristics associated with the concepts have been emphasized. Moreover, no studies have focused on the higher order thinking skills involved in analyzing and explicitly identifying the similarities and differences between concepts on the basis of their characteristics. Finally, no studies have focused on the higher order thinking skills involved in creating categories or names for the similarities and differences between two concepts. All of these are important skills if students are to be able to speak and write about the comparisons that they make.

Thus, the purpose of the current research was to develop and test the effectiveness of an instructional routine combined with a graphic organizer for the live instruction of comparisons between and among complex concepts found in secondary curricula. A major goal of the research was to develop a routine that could enhance the performance of all types of students, including high-achieving (HA), normal-achieving (NA), and low-achieving (LA) students as well as students with LD who had been included in classes according to the Individuals with Disabilities Education Act (U.S. Department of Education, 1997). Study 1 was conducted in an experimental setting to show whether the routine had positive effects under tightly controlled conditions. In Study 2, the ability of 10 general education teachers to develop graphic comparison organizers and to implement the instructional routine in their secondary science and social studies classrooms was determined. Descriptive data regarding the types and numbers of concepts that teachers selected for comparisons before and after learning the instructional routine were also collected.

The Concept Comparison Device and Routine

The focus of these studies was a graphic device called the *Concept Comparison Table* and an associated instructional routine called the *Concept Comparison Routine* (Bulgren, Lenz, Deshler, & Schumaker, 1995). The device and routine were based on teaching procedures that embodied the philosophy of C*ontent Enhancement* (Bulgren & Lenz, 1996; Lenz, Bulgren, & Hudson, 1990; Schumaker et al., 1991), which is a process of teaching scientific or cultural knowledge to students of diverse ability in which (a) both group and individual needs are met; (b) the integrity of the content is maintained; (c) critical features of the content are selected, organized, and transformed in a manner that promotes effective and efficient information processing; and (d) the instruction is carried out in a partnership with students. The device and

the routine were developed through an interactive process between a group of secondary teachers and the researchers. (For more on this process, see Lenz, Schumaker, & Deshler, 1991.)

The Concept Comparison Table and Concept Comparison Routine were designed to be incorporated into normal content instruction by a teacher to enhance students' understanding of items within sets of conceptual information and the similarities or differences between or among those items. The Concept Comparison Table (see Figure 1 for an example) is a one-page teaching tool used to visually display information about one comparison. A

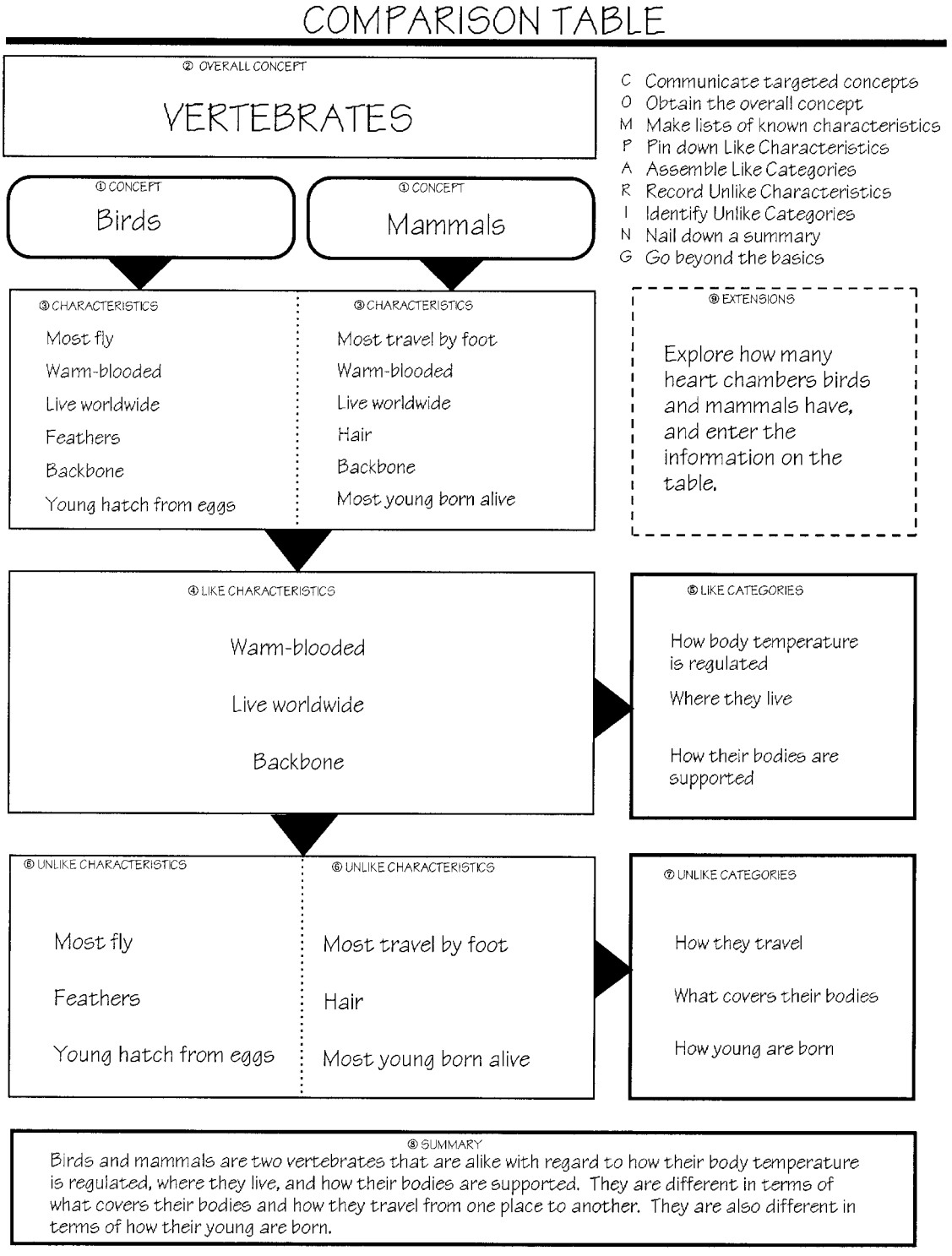

Figure 1. Sample Concept Comparison Table for the concepts *birds* and *mammals*.

comparison was defined as an examination of two or more items to identify characteristics that are alike and those that are different (Bulgren et al., 1995). The table was designed so that either concepts, that is, categories into which ideas, events, or objects can be grouped, or examples of concepts, that is, representatives of a concept category, could be displayed within it. For example, two concepts are mammal and bird. Examples of mammals are humans and whales, and examples of birds are pelicans and cardinals. Thus, within a single Concept Comparison Table, democracy and communism (two concepts) can be compared; alternatively, the government of China and the government of France (two examples) can be compared. Hereafter, such terms selected to be compared in this study will be referred to as *items of conceptual information* regardless of whether they are concepts or examples of concepts.

Further, the Concept Comparison Table was designed so that it could be used in various content areas. Regardless of the content area, graphic shapes, numbers, and word cues on the Concept Comparison Table guide the student and teacher in exploring two or more items of conceptual information. For example, in Figure 1, the concepts to be compared are *birds* and *mammals*, which are subsumed under a larger concept, *vertebrates*. Space is provided under each of the concept names for listing their characteristics. Below these spaces is a space for identifying the characteristics that are alike across the concepts (like characteristics). To the right of this space is an additional space for naming the larger category group (like categories) associated with each like characteristic. Below the box labeled *like characteristics* are spaces for identifying the characteristics that are different across the two concepts (unlike characteristics) and for naming the larger category group associated with each pair of different characteristics (unlike categories). Space is provided at the bottom of the table for synthesizing an understanding of the similarities and differences between or among items. Finally, space is provided in the upper right portion of the table for presenting a challenge to students that requires them to extend their understanding of the similarities and differences between or among the items of conceptual information. This space is labeled *extensions.*

Thus, the purpose of the Concept Comparison Table is to display information related to the similarities and differences between two or more items of conceptual information. Its graphic forms are set up in a sequence that can be logically followed to analyze those similarities and differences and to summarize them in a concise statement.

The Concept Comparison Routine, the instructional routine used to create the Concept Comparison Table with the students, contains three instructional phases: *Cue, Do,* and *Review*. The overall goal associated with these phases is that the Concept Comparison Table be developed on an interactive basis between the teacher and the students to help the students understand the information. Although some parts of the routine are directly under teacher control (e.g., announcing the items of conceptual information to be compared), students are to be involved interactively in a learning partnership whenever possible as the table is completed.

During the Cue Phase, the conceptual information to be compared is named. Then, the students are cued about the importance of understanding the information, they are instructed to take notes, and they are informed that the teacher will be helping them understand the conceptual information through the use of a Concept Comparison Table. Thus, the purpose of this phase is to get students ready to construct the table.

The Do Phase of the Comparison Routine involves the construction of the Concept Comparison Table by the students, as much as possible, with teacher guidance. Through the use of a series of steps in this phase, the teacher leads the students through the logical cognitive process involved in making a comparison and creating a summary statement about that comparison. The instructional sequence follows the numbers in the sections on the Concept Comparison Table (see Figure 1) and includes the following: (a) naming the items of conceptual information to be compared, (b) naming the overall concept category into which the conceptual items to be compared fit, (c) examining characteristics of each item, (d) identifying the characteristics that are alike, (e) identifying the categories associated with the like characteristics, (f) identifying the characteristics that are different, (g) identifying the categories associated with the different characteristics, (h) writing the summary statement, and (i) responding to a challenge that requires extending the students' understanding of the conceptual information into new areas.

The Review Phase involves reviewing the information in the table, checking students' understanding of the information, and discussing the process involved in analyzing conceptual information to make meaningful comparisons. The purpose of this phase is to ensure that students not only can speak about the information in the table but also can explain the cognitive processes involved in analyzing a comparison and how those cognitive processes can be applied to other comparison tasks.

Study 1: Effects of the Routine in Specially Designed Lessons

Method

Settings

The study took place in classrooms in one high school (serving approximately 1,650 students) and two middle schools (serving approximately 950 and 700 students, respectively) in the same suburban midwestern school district in eastern Kansas. All of the classrooms had desks, chairs, and an overhead projector and screen.

Participants

A total of 107 students enrolled in 7th-, 8th-, 10th-, 11th-, and 12th-grade science classes participated. They volunteered for the study by returning consent forms signed by themselves and their parents. Within designated subgroups, they were randomly assigned to participate in either an experimental or a control group by their teachers. Subgroups within each class included the following: HA, NA, LA, and students with LD. HA students were those who had received no more than one grade below the A or B level in academic courses in either semester of the current school year and had a grade point average (GPA) of 3.5 or higher on a 4.0 scale. NA students were those who had received no more than one grade below the C level in academic courses in either semester of the current school year and had a GPA below 3.5. LA students were those who had received at least two grades below the C level in academic courses during at least one of the two semesters of the current school year. Students with LD had been classified as such by their school district following district and state

guidelines for identifying students as having LD. Fifty-five students were in the experimental group: 12 HA, 17 NA, 6 LA, and 20 LD. Fifty-two students were in the control group: 9 HA, 16 NA, 10 LA, and 17 LD. Therefore, there were a total of 21 HA students, 33 NA students, 16 LA students, and 37 students with LD students (see Table 1).

The students with LD in the experimental group had a mean full-scale Wechsler Intelligence Scales for Children—Revised (WISC–R; Wechsler, 1974) percentile score of 33.72 and a mean Woodcock–Johnson Psychoeducational Battery (W–J; Woodcock & Johnson, 1977) Reading percentile score of 20.98. The students with LD in the control group had a mean full-scale WISC–R percentile score of 37.73 and a mean W–J Reading percentile score of 22.82.[1]

The Content Lessons

A topic, tropical diseases, was selected for the lesson to be presented. This topic was selected because (a) a lesson could be designed about the topic that contained information of the type students in a science content classroom might be expected to understand, (b) the science teachers agreed that information associated with the topic was valid content to be presented in their classes, (c) it was a topic of importance worldwide, and (d) the teachers concurred that the topic had not been covered in their courses and predicted that their students would have limited prior knowledge about it.

After notes for a lecture on the topic had been prepared, information relating to conceptual information about two tropical diseases (i.e., malaria and snail fever) was identified by the researchers. This level of conceptual information was selected to parallel the types of information teachers indicated that they would select for their content instruction. Two lecture–discussion scripts were then designed: one for the experimental group and one for the control group. The scripts were identical in certain respects and different in others. For example, embedded in both scripts were identical statements to be made by the presenters about each of the diseases and to be displayed on an overhead transparency to facilitate note-taking and understanding. The exact words to be written on the overhead transparency appeared in boldface print throughout the scripts as prompts to the presenter. Both scripts contained the same presentation of logically ordered information on the two tropical diseases. In addition, at the beginning of both scripts, the topic was introduced, and statements were included to cue the students about the importance of the information and to prompt the students to take notes. The two scripts also contained the same prompts for eliciting student involvement during the instruction.

The scripts were different in the portion of the presentation during which the items targeted for comparison (malaria and snail fever) were discussed. In this portion of the script for the experimental group, the information about the diseases and the similarities and differences between them followed the format of the Concept Comparison Routine. In the script designed for the control group, the same information about the two diseases was taught using a traditional lecture–discussion format.

The scripts were presented to four judges, who were asked to determine the content similarity of the two scripts. Of the four judges, one held a bachelor's degree and had over 60 hr of graduate study and extensive experience in teaching and curriculum. The second held a master's degree in biology and had 36 hr of graduate study in secondary science education. The third judge held a doctorate (PhD) in special education and was certified to teach at the secondary level. The fourth judge had expertise in test construction and assessment and was pursuing a graduate degree in educational psychology and research. All agreed that the same information was presented in both scripts.

Table 1
Experimental Design and Number of Students in Each Group

Group	HA	NA	LA	LD
Control				
1A[a]	0	0	1	4
3A[b]	4	4	1	2
5A[a]	0	5	1	1
2B[a]	3	2	2	2
4B[b]	2	3	2	2
6B[b]	0	2	3	6
Experimental				
2A[b]	3	5	0	2
4A[a]	0	4	2	5
6A[a]	1	2	0	5
1B[b]	2	2	0	4
3B[a]	5	1	1	2
5B[b]	1	3	3	2

Note. *A* groups were instructed by Researcher A; *B* groups were instructed by Researcher B. HA = high achieving; NA = normal achieving; LA = low achieving; LD = learning disabled.
[a] Another room. [b] Regularly assigned room.

Measurement Systems

Student test. From the information contained in the scripts, a test was designed to assess both recall of information and recognition of information. For the recall part of the test, students were asked to think about how malaria and snail fever were the same. Space was provided to list five characteristics that were the same for both diseases (e.g., both cause weakness). Next, students were asked to name the higher level category associated with each characteristic (e.g., weakness can be categorized as a symptom of the diseases). Then, students were asked to think about how malaria and snail fever were different. Five spaces were provided for students to write characteristics that were found in malaria but not found in snail fever (e.g., malaria is carried by a mosquito), and five spaces were provided for students to write characteristics that were found in snail fever but not found in malaria (e.g., snail fever is carried by a snail). Five spaces were then provided for the student to list categories associated with the pairs of different characteristics (e.g., the carriers of both diseases can be categorized as secondary hosts).

This recall portion of the test was scored in two ways. The first score reflected the number of individual characteristics or categories recalled by the students. One point was awarded for the correct name of each characteristic or category up to a total of 25 possible points. The percentage of items recalled is referred to as the *Recall Score*.

The second score reflected the students' ability to recall complete sets of characteristics and categories. One point was awarded if the student had written the complete set containing both the characteristic(s) and the category containing each characteristic that was alike or different, and zero points were awarded if either the characteristic(s) or the name of the category containing that characteristic(s) was wrong or missing. A total of 10 points was available. The percentage of complete sets recalled is referred to as *Complete Set Score.*

For the recognition part of the test, a 75-item assessment device was constructed to measure students' recognition of characteristics associated with each concept. Students were presented with a list of 25 words or phrases (e.g., *virus, microscopic, inch-long*), written in a column on the left side of a sheet of paper. Three columns of blanks were arranged to the right of each word or phrase. The first column was labeled *malaria,* the second column was labeled *snail fever,* and the third column was labeled *neither.* If the word or phrase was associated with malaria, students were to place a check mark in the first column. If the word or phrase was associated with

[1] A complete table of demographic information is available from Janis A. Bulgren.

snail fever, they were to place a check mark in the second column. If the word or phrase was associated with both malaria and snail fever, they were to place a check mark in the blanks in both the first and second columns. If the word or phrase was associated with neither malaria nor snail fever, they were to place a check mark in the third column. One point was awarded for the correct placement of each check mark or for the correct absence of a check mark for a total of 75 possible points. The percentage of answers correct on this portion of the test was referred to as the *Recognition Score.*

Finally, a total score was compiled by combining the points received for the Recall, Complete Set, and Recognition Scores. This was done to represent the percentage of total possible points earned by the students (110 points), a score similar to one that students might normally receive in a content classroom on a test comprising several different types of measures. Points awarded for correct responses were added for all three scores and divided by 110 to arrive at the total percentage of correct responses.

Interscorer reliability. Interscorer reliability was determined by having two scorers independently score a random sample (15%) of the tests. The two observers' recordings were compared item by item. For items in the section designed to elicit recall of individual characteristics and categories, scorers agreed 567 times out of 625 opportunities to agree (range 80% to 100%) for 90.7% agreement. For items in the section designed to elicit recall of complete sets of characteristics and categories, scorers agreed 222 times out of 250 opportunities to agree (range 70% to 100%) for 88.8% agreement. For items in the recognition section, scorers agreed 1,873 times out of 1,875 opportunities to agree (range 98.7% to 100%) for 99.8% agreement. Overall, scorers agreed 2,662 times out of 2,750 chances to agree for 96.8% agreement.

Procedure

At the beginning of the appointed class period (which was a regularly scheduled class time), two researchers visited each participating science class. The half of the students who were selected to participate in the experimental group remained in the classroom with one of the researchers, and the other half of the students comprising the other group went to another classroom with the other researcher (or visa versa, according to the experimental design). The classroom teachers had randomly assigned the students to groups in such a way that students from each subgroup (HA, NA, LA, and LD) were approximately equally present in each group. At the beginning of the lesson, all students were provided a pencil and a blank piece of paper on which to take notes. The experimental group was taught the lesson comparing the two tropical diseases (malaria and snail fever) through the use of the Concept Comparison Table and Routine by one of the researchers. The control group was taught the lesson using the traditional lecture–discussion format. The researchers followed the appropriate script for each lesson word for word. They spent the same amount of time on the lesson (40 min). Student notes were collected at the end of the lesson.

On the next day, the test was administered to all students. Students were allowed to study from their own notes for 5 min at the beginning of the period. Test instructions were read to the students; test items were not read to the students. Students were administered the recall portion of the test at the outset of the examination period. Then that portion of the test was collected before the recognition portion of the test was passed out to students. Students were given a total of 40 min to complete the test. They were not allowed to consult their notes or each other as they took the test.

The instructors included Janis A. Bulgren, who holds a PhD in special education and has been certified to teach at the secondary level, and a research assistant, who is certified to teach and has had experience teaching at the elementary and secondary levels. Each lesson was tape-recorded to ascertain that each instructor adhered to the script. A third researcher listened to the tapes and compared the information presented by each instructor to the prepared script. The third researcher determined that the scripts were followed 100% of the time.

Experimental Design

The experimental design and number of participants are depicted in Table 1. As described above in the *Participants* section, approximately half of each of the subgroups in each of the six participating science classes was randomly selected by their regular classroom teachers for the experimental group, and half was selected for the control group. (Unmatched numbers are the result of uneven numbers of students in a subgroup.) Two major variables were controlled with this design: the instructor and the classroom. For half of the classes, Researcher A was the instructor for the experimental group, and Researcher B was the instructor for the control group. For the other half of the classes, Researcher B was the instructor for the experimental group, and Researcher A was the instructor for the control group. For half of the experimental groups, the regularly assigned classroom was the setting; for the other half of the experimental groups, another classroom was the setting. Likewise, for half of the control groups, the regularly assigned classroom was the setting; for the other half of the control groups, another classroom was the setting.

Therefore, in this design, the classrooms and treatments were crossed rather than having students nested within classrooms that were then nested within treatments. Furthermore, the treatment was implemented by two different researchers, thoroughly trained in the treatment protocol and counterbalanced in their assignment to treatment and control groups. This procedure eliminated nesting in this part of the design. Therefore, the usual reasons to use classrooms as the unit of analysis (e.g., that the research question involved some effect at the classroom level or because students or teachers were nested in classrooms) were not relevant.

Results

Mean scores for each of the subgroups within the experimental and control groups are shown for the following measures in the following figures: the Recall Score in Figure 2, the Complete Set Score in Figure 3, the Recognition Score in Figure 4, and the Total Score in Figure 5.

A multivariate analysis of variance was conducted with two between-subjects factors and three outcome variables. The between-subjects factors were student type (LD, LA, NA, and HA) and experimental condition (control and experimental). The three outcome variables were the three scores: the Recall Score, the Complete Set Score, and the Recognition Score. Because the research questions did not concern a comparison of the outcome scores with each other, they were not considered as a factor in the analysis. The value of the F statistic for Wilks's lambda, the probability for the statistic, and a measure of the effect size were determined. The effect size reported is the η^2 statistic, which may be interpreted according to the following guidelines given by Cohen (1988): $\eta^2 = .010$ is small, $\eta^2 = .059$ is medium, and $\eta^2 = .138$ is large. The criterion alpha level used for statistical significance was .05.

The multivariate analysis showed no significant interaction between student type and experimental condition. Statistically significant differences were found for experimental condition, $F(3, 97) = 6.91$, $p < .001$, $\eta^2 = .176$, and student type, $F(9, 236.22) = 4.85$, $p < .000$, $\eta^2 = .129$.

Because the primary focus of this research was the effect of the teaching method (experimental or control condition), follow-up analyses concentrated on the effects of the teaching method for the

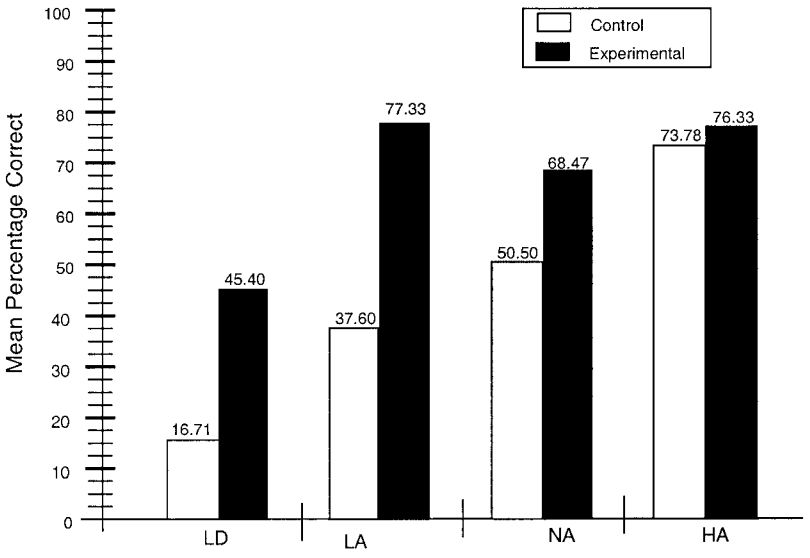

Figure 2. Mean percentage scores earned by students in experimental and control conditions on test designed to assess recall of individual characteristics and categories (Recall Score). LD = learning disability; LA = low achieving; NA = normal achieving; HA = high achieving.

different outcome measures and within the different student types. Differences between student types were expected and were not included in the research questions for this study; hence, no follow-up tests were done with the different student types.

For each of the three outcome measures, significant differences were found between the experimental group and the control group. For the Complete Set Score, $F(1, 99) = 20.53$, $p < .001$, $\eta^2 = .172$. For the Recall Score, $F(1, 99) = 16.70$, $p < .001$, $\eta^2 = .144$, and for the Recognition Score, $F(1, 99) = 7.41$, $p < .01$, $\eta^2 = .070$.

Additional follow-up analyses of experimental versus control group were conducted within each student type. The results for each student type are reported below.

The multivariate analysis using all three scores revealed that the experimental students with LD performed significantly better than the control students with LD, $F(3, 33) = 3.46$, $p = .027$, $\eta^2 = .239$. Univariate follow-up analyses for each of the three outcome variables for students with LD showed that the Concept Comparison group scored significantly better than the control group on each of the three tests: the Complete Set Score, $F(1, 35) = 10.24$,

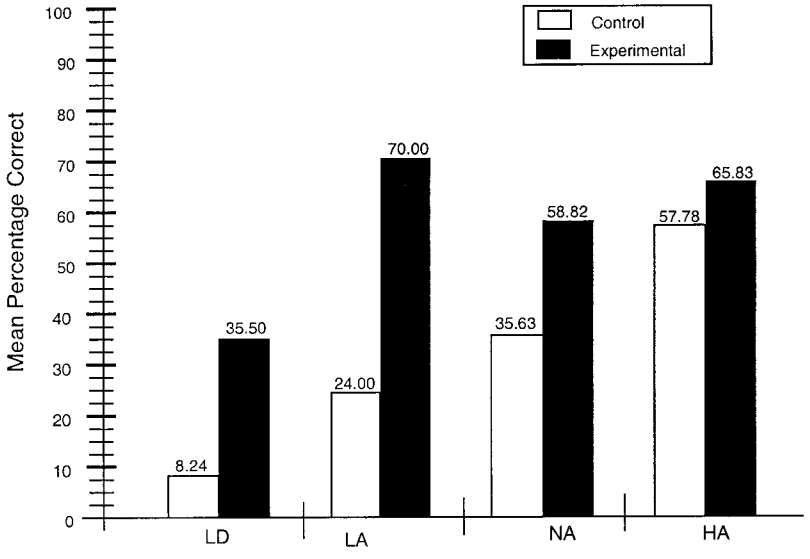

Figure 3. Mean percentage scores earned by students in experimental and control conditions on test designed to assess recall of complete sets of characteristics and categories (Complete Set Score). LD = learning disability; LA = low achieving; NA = normal achieving; HA = high achieving.

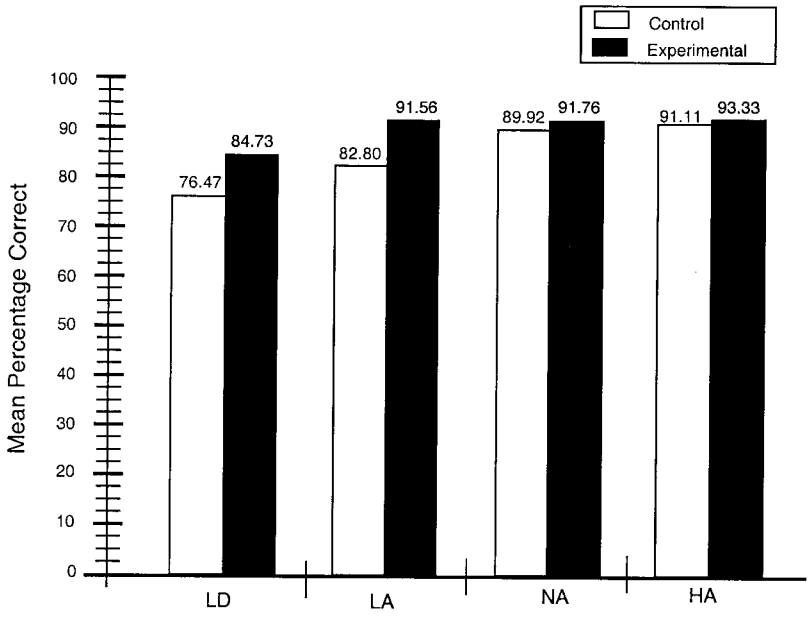

Figure 4. Mean percentage scores earned by students in control and experimental conditions on test designed to assess recognition of characteristics (Recognition Score). LD = learning disability; LA = low achieving; NA = normal achieving; HA = high achieving.

$p = .003$, $\eta^2 = .226$; the Recall Score, $F(1, 35) = 10.88$, $p = .002$, $\eta^2 = .237$; and the Recognition Score, $F(1, 35) = 4.84$, $p = .035$, $\eta^2 = .121$.

For the LA students, the follow-up analyses comparing the students who participated in the Concept Comparison Routine with the students who were presented the same information in a traditional lecture–discussion format yielded the following results: For the Complete Set Score, $F(1, 14) = 8.39$, $p = .012$, $\eta^2 = .375$; for the Recall Score, $F(1, 14) = 6.41$, $p = .024$, $\eta^2 = .314$; and for the Recognition Score, $F(1, 14) = 2.44$, $p = .14$, $\eta^2 = .148$. The multivariate analysis using all scores was not statistically significant, $F(3, 12) = 2.55$, $p = .105$, $\eta^2 = .389$. The very large effect

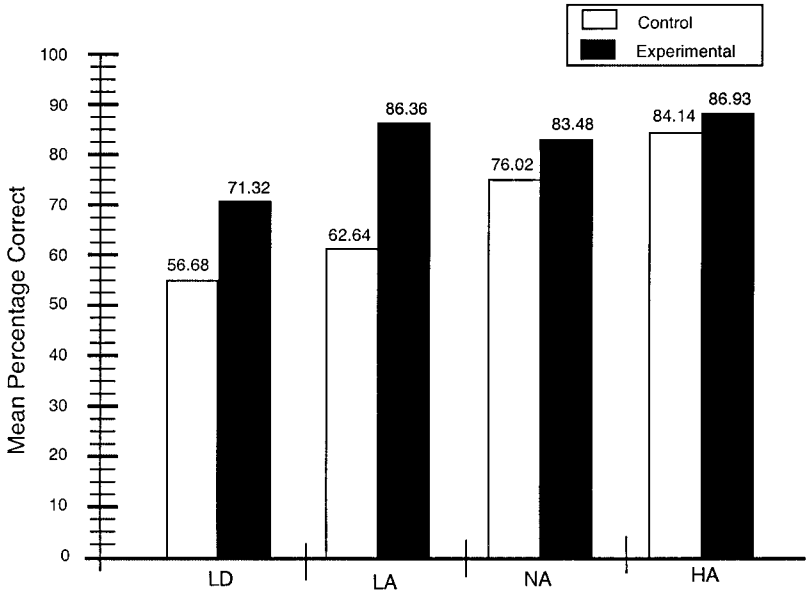

Figure 5. Mean percentage combined scores earned by students in control and experimental conditions on combined tests (Total Score). LD = learning disability; LA = low achieving; NA = normal achieving; HA = high achieving.

size indicated that the lack of a statistically significant difference was likely due to low power (observed power = .485), which was probably a consequence of having only 6 students in the experimental group and only 10 students in the control group. With such a small number of students in each group, one must be cautious in the interpretation of results; however, the differences between the experimental and control groups on two of the three outcome measures showed very large effect sizes and small probabilities. For the third measure, there was also a large effect size, but only a moderately small probability.

The multivariate analysis for the NA students also indicated a relatively large effect size (.187) but again relatively low power (observed power = .505) to detect differences, $F(3, 29) = 2.21$, $p = .107$, $\eta^2 = .187$. In this case, the experimental and control groups had 17 and 16 students, respectively. As with the LA students, NA students who participated in the Concept Comparison Routine earned significantly higher Complete Set Scores than NA students who were presented the same information in a traditional lecture–discussion format, $F(1, 31) = 6.32$, $p = .017$, $\eta^2 = .169$. The effect size for the differences between the groups on the Recall Score was between medium and large, $F(1, 31) = 4.06$, $p = .053$, $\eta^2 = .116$. There was no statistically significant difference between the Recognition Scores earned by the NA experimental group and the NA control group. There also were no statistically significant differences and no large effect sizes between the HA experimental and control groups on any of the three outcome measures.

Further analysis was done to determine levels of student performance on the tests when judged by standards that are often applied to content test performance in secondary general education classes. That is, results were analyzed to determine the percentage of students who would have performed at a level generally deemed as *passing* (i.e., by earning a score of 60% or above). Students' Total Scores represented passing grades (i.e., scores above 60%) for the following percentages of students: for students with LD, 29.41% in the control group and 70.76% in the experimental group; for LA students, 50.00% in the control group and 83.33% in the experimental group; for the NA students, 87.50% in the control group and 94.12% in the experimental group; for the HA students, 100% in the control group and 91.65% in the experimental group.

Once researchers ascertained that the use of the Concept Comparison Table and Routine could produce significantly better results compared with a traditional lecture–discussion format in a controlled situation, the next step was to explore whether teachers of inclusive general education classes could incorporate the researched technique into regular classroom practice. Therefore, Study 2 was conducted to determine the effects of training in the routine on teacher use.

Study 2: Effects of Training on Teacher Use of the Routine

The purpose of Study 2 was to explore (a) teacher response to a professional development session related to the Concept Comparison Table and Routine in terms of the instruction they delivered in their classes, (b) the numbers and types of conceptual information sets that were selected for comparison by the teachers before and after they were introduced to the Concept Comparison Table and Concept Comparison Routine, and (c) teacher and student satisfaction with the use of the routine.

Method

Settings and Participants

Participants in Study 2 taught in two school districts located in suburban areas of eastern Kansas. Ten general education secondary content teachers who taught inclusive classes volunteered to participate in the study after being approached individually. One high school social studies teacher in School District A (school serves approximately 1,700 students) taught American history to students in the 11th grade. In another high school in the same district (school also serves approximately 1,700 students), one science teacher taught a 9th-grade class called Introduction to Investigative Science Skills, and another teacher taught 9th-grade world geography. In a middle school in the same district (school serves approximately 350 students), 1 teacher taught American history to 8th-grade students, 1 taught life science to 8th-grade students, and 1 taught science to 7th-grade students.

In School District B, 1 teacher (school serves approximately 1,750 students) taught basic biology to students in the 10th and 11th grades, and 2 other teachers taught geography in a junior high (school serves approximately 550 students) to 7th-grade students. Another teacher in that district taught American history to students in the 8th grade in another junior high school (school serves approximately 650 students). All teachers (7 White women and 3 White men) were offered $80 each to participate. This fee compensated the teachers for the time they spent in the professional development session. Students in the classes of these teachers were administered a satisfaction questionnaire.

Measurement Systems

Comparison content sheet. A measurement system in the form of a recording sheet and checklist was constructed on which were written all of the pairs of information items (e.g., mammals and birds) that the teachers had selected for presentation in their classes. After the names of each pair or set, space was provided to place a check mark if a judge identified the pair or set as concepts; another space was provided to place a check mark if a judge identified the pair or set as examples of a concept. In addition, spaces were provided to indicate if the items selected for comparison were abstract, concrete, or both. A *concept* was defined as a "category into which ideas, events, or objects can be grouped," and an *example* was defined as a "representative that fits within a concept category." An *abstract* concept–example was defined as "an item expressing a quality apart from an object." A *concrete* concept–example was defined as "an item having observable or demonstrable physical properties." The Comparison Content Sheets were submitted to three judges. After each judge had nominated each pair of items as either examples or concepts and had classified the pair with regard to its abstract or concrete nature, the descriptive data were compiled.

Implementation checklist. A measurement system in the form of a checklist was used to assess the level of teacher performance in implementing the Concept Comparison Routine in the classroom. It contained items designed to reflect the parts of the Concept Comparison Routine: (a) naming the general topic for the lesson, (b) cuing the students about the importance of understanding the information involving comparisons that was to follow, (c) cuing the students to take notes over the lesson and to participate in the discussion, (d) cuing students how a graphic device would be used to enhance understanding, (e) naming items to be compared, (f) exploring the characteristics of each item, (g) identifying characteristics that were alike, (h) identifying the larger categories for the characteristics

that were alike, (i) exploring the characteristics that were different, (j) identifying the larger category for the characteristics that were different, (k) creating a summary statement about the items and how they were alike and different, (l) extending exploration of the comparison, and (m) reviewing the content and the process of using of the Concept Comparison Table.

Each item on the checklist was objectively defined in an Evaluation Manual designed to guide observers as they watched teachers teach their classes and use the Concept Comparison Routine. Included in the Evaluation Manual were the following: (a) a discussion of possible ways to use the routine within a lesson, (b) a description of the Implementation Checklist, (c) instructions regarding how to score elements involved in the use of the routine, (d) examples and nonexamples of each element of the routine, and (e) the points to be awarded for each element in the routine.

In all observations, observers were instructed to begin recording information on the Implementation Checklist about the presentation of a comparison when any of the following occurred: (a) The teacher cued students about the importance of understanding similarities or differences in content information, (b) the teacher cued the students about the use of a graphic device to help them understand a comparison and explained the understanding to be gained, (c) the teacher presented or elicited characteristics of two or more items, (d) the teacher presented or elicited characteristics shared by two or more items, or (e) the teacher presented or elicited the differences between two or more items.

A point value was assigned to each item on the Implementation Checklist. A total of 100 points was possible, and each item was assigned points ranging from 5 to 15. For example, 5 points were awarded when the teacher cued the students about the importance of understanding the information. The mastery criterion was arbitrarily set at 85 points.

Teacher and student satisfaction. Social validity questionnaires were designed to determine student and teacher satisfaction with the routine. First, an instrument was devised to determine teacher satisfaction with the Comparison Table and the associated teaching routine. There were 20 items on the questionnaire; each item was in the form of a 7-point Likert-type scale ranging from 1 (*completely dissatisfied*) to 7 (*completely satisfied*). Specific items explored flexibility, ease of use, acceptability of preparation time, whether students with and without LD could learn better when the Concept Comparison Table was used, whether the students with and without LD perceived the tables as helpful, whether achievement for students with and without LD improved when the table was used, whether attention or time on task increased for students with and without LD, whether note-taking skills improved for students with and without LD as a result of teacher use of the table and routine, whether study time increased for students with and without LD, and whether students with and without LD used the table in studying for tests. In addition, data were collected regarding the likelihood that teachers would continue using the table and routine in their classes, whether they would recommend them to other teachers, and whether they would recommend them to other teachers if in-service instruction was available.

Similar information was also elicited from the students through administration of a seven-item questionnaire. Again, each item comprised a 7-point Likert-type scale. Particular items determined how satisfied the students were that the Comparison Table helped them to follow what the teacher was saying, to take notes, to focus attention on what was important, to study for tests, to do well on tests, and to improve their grades, as well as how satisfied they were with the new way of teaching compared with when the teacher did not use the Concept Comparison Table and Routine.

Interscorer reliability. Interscorer reliability was determined on the Implementation Checklist and the Comparison Content Sheets by having two scorers independently score 15% of the presentations (i.e., 12 of 81 observations). The points awarded by the two observers were compared item by item for each pair of checklists and each pair of sheets. The percentage of agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. For the Implementation Checklist, the scorers agreed 153 times out of 168 opportunities to agree (total percentage of agreement = 91.1%; range = 78.6% to 100%). For the Comparison Content Sheets, the scorers agreed 100% of the time.

Procedure

Baseline. Baseline data were collected before the professional development session. Teachers were requested to inform researchers when they would be teaching content that involved comparisons that the students would be expected to understand. They were then observed during at least three class sessions that they had identified. The observer(s) began recording information on the Implementation Checklist any time when the teacher named conceptual information or examples to be compared, elicited or presented items to be compared in a visual format, specified a way that students could understand similarities and differences, or presented information to be compared in some way. The Comparison Content Sheets were completed after the presentation.

Teacher instruction. Procedures were developed for teachers to use in preparing Concept Comparison Tables and delivering instruction using the Concept Comparison Routine. These procedures were incorporated into a teacher's manual that was used in teacher training and provided to the teachers for use as a reference guide. The manual included rationales for developing comparisons between or among conceptual items in content-area instruction, a definition of a comparison, examples of information that could be taught using the table and routine, a description of the table, guidelines for preparing tables, examples of tables, explanations about how to introduce the table to students, a step-by-step description of how to use the routine, guidelines for evaluating how well students have learned the conceptual information taught, and planning forms for use in the development of a table.

Researchers presented information to the teachers in a 2-hr workshop session that included a description of the Concept Comparison Table and the presentation routine associated with the table, the rationales for using the Concept Comparison Table and Routine to present content information, a demonstration by researchers of the use of the table and routine, practice in developing the tables, and practice presenting the routine. Researchers provided clear explanations to the teachers that good implementation of the routine included all the points discussed (see the section on the *Implementation Checklist* above for the listed and discussed items) regarding cuing the students about the importance of the routine and the information, interactively developing the Comparison Table, and carefully reviewing the content and the process by which it was learned.

Implementation. The teachers selected conceptual information involving comparisons that they judged important for the students to understand. After the conceptual information had been selected from the subject matter, each teacher prepared the Concept Comparison Tables and planned how the information would be presented. Researchers consulted with the teachers individually as requested by the teachers; this usually consisted of brief discussions between a teacher and a researcher after a teacher's implementation of the Concept Comparison Routine.

After the teachers had planned the instruction, they implemented the routine in their classes according to their own schedules. They informed the researchers as to what days they would be implementing the routine. An observer(s) attended the specified classes and recorded information on the Implementation Checklist and on the Comparison Content Sheet. As a measure of the routine's social validity, the students and teachers were administered the satisfaction questionnaires at the end of the school year within 1 week of the teachers' last use of the comparison routine.

Experimental Design

A multiple-probe across-subjects design (Horner & Baer, 1978), a variation of the multiple-baseline design (Baer, Wolf, & Risley, 1968) was

used. Two teachers participated in each design; thus, there were five iterations of the design. Each teacher was observed at least three times during baseline. Five teachers whose baseline data (according to the Implementation Checklist) were stable after three observations received training. After training, they were observed several more times. Meanwhile, the remaining 5 teachers were observed under baseline conditions during at least one more lesson. Once the first 5 teachers had shown that they could implement the routine and when the remaining teachers' baseline observations were stable, the second set of 5 teachers received training. Observations for all of the teachers continued throughout the rest of the semester when they indicated that they would be teaching a comparison lesson.

Results

Teacher Performance of the Routine

During the 42 baseline class periods, which the teachers had identified as class periods where information was to be compared, the teachers provided a way to understand the similarities and differences between or among items of information only 10 times (or 23.8% of the classes). In fact, only 4 of the 10 teachers presented ways to understand similarities and differences between or among items of information during baseline, and 2 teachers accounted for 6 of those 10 presentations. During baseline, even when the unit or topic selected directly involved a comparison, the instruction often did not focus on the comparison. For example, there was no instruction on a comparison even when the unit title was Minerals and Rocks or Simple and Complex Machines. Instead, the items were dealt with sequentially; that is, one item was described, and then the other item was described. After training, the teachers provided ways to understand similarities and differences between or among items of content information in 39 of the 39 observations, or in 100% of the observations.

Figures 6, 7, 8, and 9 show the teachers' performances with regard to implementation of the Concept Comparison Routine, as reflected by their percentage scores on the Implementation Checklist. In Figure 6, the performance of 4 teachers is shown, and in Figures 7, 8, and 9, the performance of 2 teachers is shown on each. For each teacher, baseline performance is shown to the left of the vertical line in each graph, and posttraining performance is shown to the right of this line.

During the 42 baseline observations, the teachers' scores on the checklist ranged from 0% to 50% (M = 8.5%). In general, before training, the behaviors for which teachers earned points on the checklist were those associated with identifying the items for comparison and eliciting or presenting the characteristics of each.

After training, the teachers reached or exceeded the 85% mastery level in 38 out of the 39 classes observed. The average teacher score after training was 92.8% (range = 60% to 100%). The teachers presented only one instance of a comparison during each session. After Teacher 5 failed to reach mastery in one instance, she consulted with the researchers and then exceeded the mastery level during the next four presentations, achieving scores above 90%. In two of these lessons, Teacher 5 used an alternate procedure in which she interacted with students as they created their own concept Comparison Tables. Points were awarded for the class as if the teacher had written the information on the board herself. These were the last two classes observed for Teacher 5,

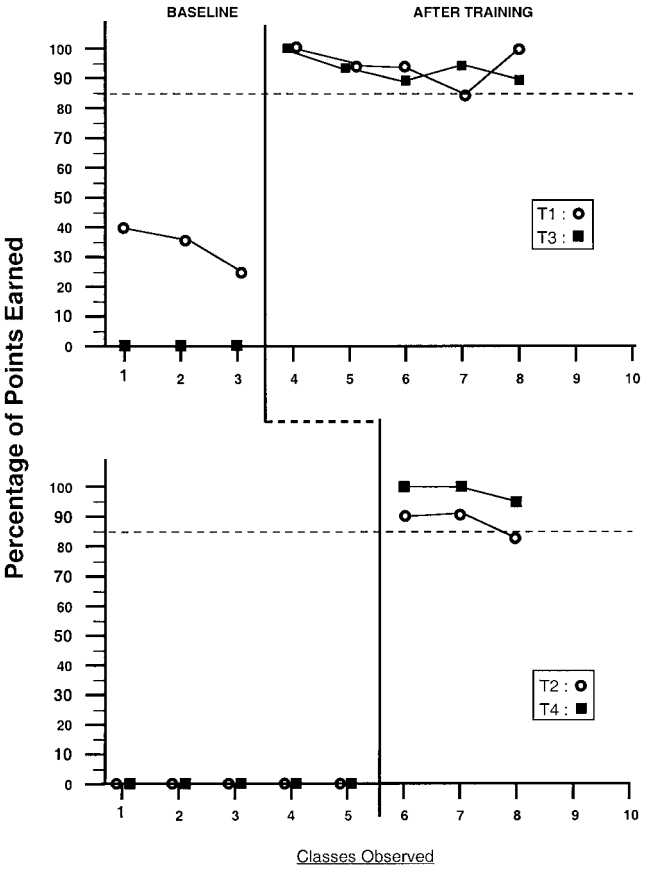

Figure 6. Percentage of points earned by Teachers (T) 1–4 on the Implementation Checklist.

and the teacher informed the researchers that she felt comfortable enough with the routine to make this adaptation.

Data were collected to determine the total number of times that the teachers provided cues about the importance of understanding similarities and differences between or among conceptual items before and after training on the Concept Comparison Routine. During the 42 baseline observations, the teachers provided cues that the information they were about to present regarding similarities and differences between or among items was important to understand only once, that is, in 2.4% of the observations. After training, the teachers cued the importance of understanding similarities and differences 31 times in 39 classes, that is, in 79.50% of the classes.

Types of Comparisons Developed by the Teachers

Analysis of the comparisons developed by the teachers before and after training in the Concept Comparison Routine provided information regarding the types of items chosen by the teachers for comparison. First, items were analyzed to determine if the comparisons chosen by the teachers involved concepts or examples of concepts. In baseline, 2 of the comparisons presented by social studies teachers involved concepts (25%), and 6 involved examples (75%). After teacher training, 8 of the social studies compar-

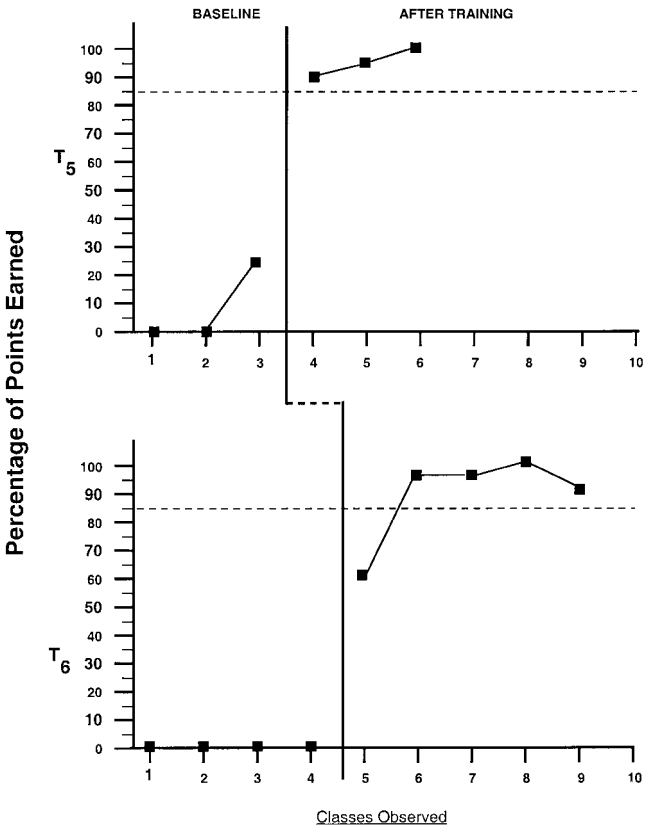

Figure 7. Percentage of points earned by Teachers (T) 5 and 6 on the Implementation Checklist.

isons involved concepts (32%), and 17 involved examples (68%). In baseline, both comparisons presented by science teachers involved concepts. After teacher training, the science teachers presented 10 comparisons involving concepts (71.43%) and 4 (28.57%) involving examples.

Second, items were analyzed to determine whether the compared concepts or examples were abstract, concrete, or both. For science, all of the items of conceptual information mentioned during baseline and in the intervention phase of the study were concrete (e.g., shield, cinder, and composite volcanoes). For social studies, all 8 comparisons in baseline involved both abstract and concrete items of conceptual information (e.g., capitalism and communism). During implementation of the routine, 22 of the 25 social studies comparisons involved both abstract and concrete items of conceptual information (88%), none involved only abstract items of conceptual information, and 3 involved concrete items of conceptual information (12%).

Third, the number of items selected for each comparison was determined. In science, during baseline, 2 items (e.g., reptiles and amphibians) were compared once (50%), and 3 items (e.g., reptiles, birds, and amphibians) were compared once (50%). After training, science teachers compared more than 2 items a total of 3 times out of 14 observations (21.43%), and 2 items were compared in the other 11 observations (78.57%). In social studies in baseline, 2 items were compared 4 times (50%), and more than 2 items were compared 4 times (50%). Five items were involved in one comparison, and 10 were involved in another in baseline. For the 25 observations during the implementation condition, social studies teachers compared 2 items 21 times (84%) and more than 2 items only 4 times (16%).

Satisfaction Questionnaires

The results of the teacher satisfaction questionnaire indicated that on average the teachers were satisfied with many aspects of the program. The mean ratings were as follows: flexibility of the routine, 6.20 (range = 4–7); ease of use, 5.90 (range = 4–7); acceptable preparation time, 6.10 (range = 5–7); that the Concept Comparison Table helped students with LD to learn facts, 5.30 (range = 3–7); that the table helped students without LD to learn facts, 6.30 (range = 5–7); that students with LD perceived the table as useful, 5.20 (range = 3–7); that students without LD perceived the table to be useful, 6.0 (range = 5–7); that achievement for students with LD improved as a result of using the table, 5.0 (range = 3–6); that achievement for students without LD improved as a result of using table, 5.67 (range = 5–6); that attention increased for students with LD, 5.80 (range = 5–7); that attention increased for students without LD, 6.10 (range = 5–7); that note-taking skills increased for students with LD, 5.0 (range = 3–6); and that note-taking skills increased for students without LD, 5.56 (range = 4–7). They were not satisfied, on average, that study time increased for students with LD, 4.50 (range = 3–6); study time increased for students without LD, 4.44 (range = 3–6); students with LD used the table in studying for tests, 4.50 (range = 2–7); and students without LD used the table in studying for tests, 4.78 (range = 2–7). The teachers responded in the following ways when they were asked to judge the likelihood that they would continue to use the table and routine in their classes, 6.20 (range = 5–7); recommend the table and routine to other teachers, 6.20 (range = 5–7); and recommend the table and routine to others if in-service instruction were available, 6.40 (range = 5–7).

The questionnaire results from 198 students indicated that on the average they were neither satisfied nor dissatisfied with many aspects of the comparison instruction. Mean ratings were as follows: (a) whether the table helped them follow what the teacher was saying, 4.96 (range = 1–7); (b) whether the table helped them to take notes, 4.78 (range = 1–7); (c) whether the table helped them to focus attention on what was important in class, 4.91 (range 1–7); (d) whether the table helped them to study for tests, 4.80 (range = 1–7); (e) whether the table helped them to do well on tests, 4.64 (range = 1–7); (f) whether they liked this new way of teaching as compared with when their teacher did not use it, 4.66 (range = 1–7); and (g) whether the table helped them to improve their grades, 4.47 (range = 1–7).

General Discussion

These studies provide additional support for the growing body of evidence that the Content Enhancement Approach to instruction (Bulgren & Lenz, 1996) can yield positive results with regard to student learning in classes comprising diverse groups of learners. They also indicate that teachers can easily learn a complex teaching routine and are satisfied with such a routine.

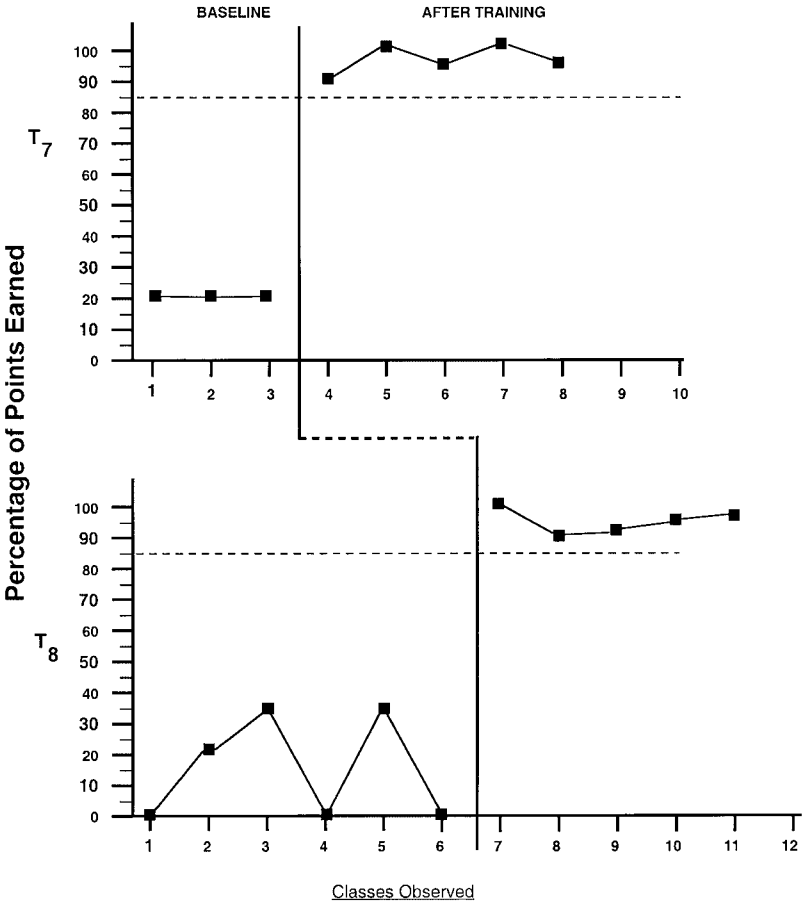

Figure 8. Percentage of points earned by Teachers (T) 7 and 8 on the Implementation Checklist.

First, Study 1 provides evidence that NA, LA, and LD students can benefit from the use of the Concept Comparison Table and Routine. Those students who seem to benefit the most are the LA and LD students. The experimental students with LD scored significantly higher than control students with LD on both recall measures and the recognition measure. The experimental LA students scored significantly higher than control LA students on both recall measures. The experimental NA students scored significantly higher than control NA students did on the recall measure that required complete sets of characteristics and categories. The effect sizes indicated that the differences between the groups were substantial. In all of the experimental NA, LA, and LD subgroups, larger numbers of students earned total scores in the passing range than in the respective control groups. Thus, Study 1 provides evidence that performance of NA, LA, and LD students on measures of higher order thinking (i.e., recall of conceptually related information) can be enhanced through the use of a graphic device and an interactive instructional routine.

Second, Study 2 provides evidence that teachers can learn to prepare a graphic comparison device and construct the device with their students quickly. The professional development session in this study was 2 hr long. All of the teachers except 1 exceeded the mastery level the first time they used the routine in their classes. The other teacher exceeded the mastery level in her second attempt after receiving feedback from a researcher. Additionally, the teachers' satisfaction ratings indicated that they were satisfied with the routine and the graphic device in many areas. Therefore, the results of the study suggest that when an instructional innovation is well defined and teachers are provided with explicit instruction and concrete examples for using it, their implementation of the innovation is likely to be successful.

Several other findings are worthy of discussion. First, in Study 1, no significant differences were found between the experimental HA students' scores and the control HA students' scores. This might have been related to a ceiling effect because the control students' mean Total Score was above 90. Like the NA students, the experimental HA students earned a substantially higher Complete Set Score than did the control HA students. However, probably because of the small number of participants in this subgroup, a significant difference was not found between the groups. Both the experimental HA students and NA students had better performance on the measure requiring the most complex type of retention of information, the Complete Set Score. This result indicates that this type of instructional routine may be most helpful to NA and HA students when information recall demands are complex.

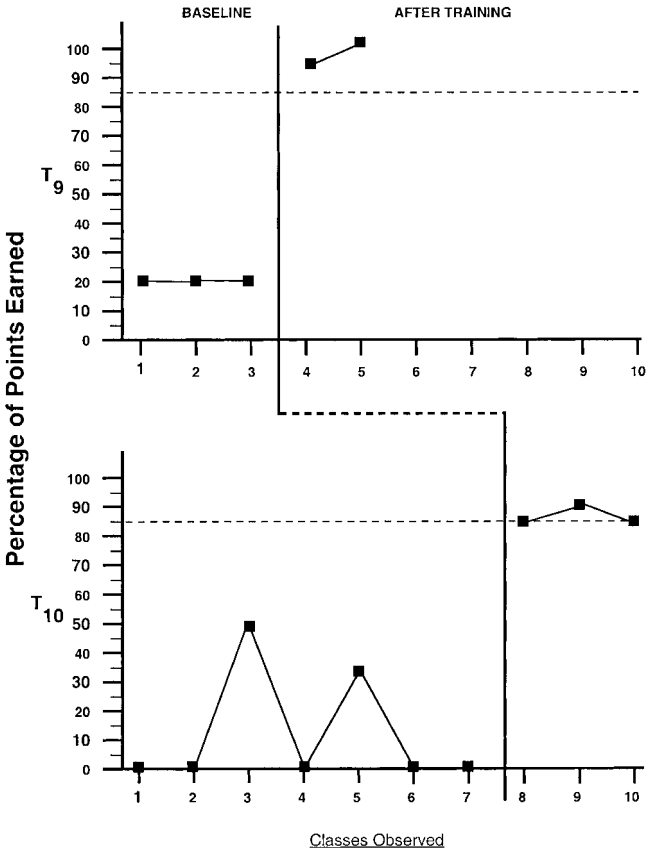

Figure 9. Percentage of points earned by Teachers (T) 9 and 10 on the Implementation Checklist.

The LA students seemed to get the biggest boost of all the subgroups with regard to learning. The mean recall scores of the LA control group were in the low failing range, and the recall scores of the LA experimental group were in the average (C) range. The mean recognition score of the LA control group was in the above-average (B) range, and the recognition score of the LA experimental group was in the superior (A) range. This indicates that use of a routine such as the Concept Comparison Routine can enhance the performance of LA students on a variety of assessments.

Although significant differences were achieved for LD students on all measures, both the experimental and control LD students performed, on average, at the failing (F) level on both recall measures during the posttest. However, the mean recognition score of the LD control group was in the average range, and the recognition score of the experimental LD students was in the above-average (B) range on the posttest. These findings may indicate that use of a routine such as the Concept Comparison Routine can enhance the performance of LD students on recognition tests to a more acceptable grade but does not bring it into the passing range on recall tests. However, because most tests solely comprise recognition items (Putnam, Deshler, & Schumaker, 1992), performance related to the recognition scores might raise the total score. Specifically, the mean total score of students in the LD experimental groups was in the average range, whereas the mean total score of students in the LD control group was in the failing range.

An interesting result of Study 2 is related to the baseline data. When teachers said that they were going to present a lesson involving a comparison during baseline, they tended to teach the information sequentially and did not explicitly make comparisons. That is, they covered the first topic and then covered the second topic. Thus, although the teachers might have been identifying comparisons as a part of their instruction, they were not focusing on those comparisons or teaching the students how to make comparisons.

Another intriguing result of Study 2 is related to the types of comparisons the teachers made. Many of their comparisons, especially those created by the social studies teachers, focused on examples of concepts and not on the higher order concepts. That is, they might choose to focus a lesson on comparing Eastern Europe and Western Europe versus on comparing democracy and communism. Why the teachers made their choices is unknown at this time. Because many of the examples chosen by the social studies teachers included both abstract and concrete characteristics, the information contained in the examples might have been so complex that it necessitated careful analysis of examples of concepts with the Concept Comparison Table. In addition, the examples might have been chosen to provide a foundation of background knowledge needed to understand the higher order concepts embedded in different types of governments and economic systems.

Additionally, the differences between the science and social studies teachers are of note. The science teachers tended to compare concepts using concrete characteristics, whereas the social studies teachers tended to compare examples of concepts using both abstract and concrete characteristics.

Of concern are some of the satisfaction ratings of the teachers and students. The teachers had mean ratings below the satisfied level when they indicated how satisfied they were with increases in study time and student use of the Comparison Table while studying for tests. Perhaps the teachers had not indicated that the table should be used to study for future tests or that some portion of course assessments would be based on information on the table. This notion was not emphasized in the teacher's manual.

All of the mean students' ratings were below the satisfied level. Why they were not pleased with the instruction is unknown at this time. Therefore, studies are needed to explore students' satisfaction with the instruction and to bolster their satisfaction, because teachers report that if students do not like an instructional procedure, they are apt to discard it (Lenz et al., 1991). Perhaps the different subgroups felt differently about the instruction, and the ratings of some subgroups pulled down the mean ratings. Certainly, some subgroups benefited more than others in Study 1. However, because the ratings were made anonymously in Study 2, the variability in ratings across subgroups is not known. Satisfaction data specific to different groups of students should be collected in future studies. Interview data would probably also be helpful in specifying student concerns.

These results extend the research on the use of graphic devices to the instruction of comparisons. Before this study, no studies had focused on the use of a graphic device related to live instruction of comparative information in secondary general education classes. These results also extend the research on the instruction of con-

cepts. Although single concepts have been successfully taught and learned through Content Enhancement Routines such as the Concept Mastery Routine (Bulgren et al., 1988) and the Concept Anchoring Routine (Bulgren et al., 2000), until the present study was conducted, no studies had focused on the comparison of two or more items of complex conceptual information. Unfortunately, no studies have investigated the effects of using several routines in combination. Therefore, future research is needed in this area. Perhaps teachers who have learned how to teach concepts using the Concept Mastery Routine or the Concept Anchoring Routine would become more comfortable with the instruction of concepts and would choose to compare concepts more often than examples.

The results of this research are limited owing to several considerations. Study 1 focused on a single comparison of two examples of tropical diseases, only one lesson was taught to the students, and the lessons were taught with close adherence to a script. In addition, both studies took place in suburban schools, and the teachers who participated in Study 2 were volunteers who were paid to participate in the professional development session. Data were not gathered on the quality of the Concept Comparison Tables, nor was the quality of tables compared across teachers. In addition, student performance data were not collected in Study 2. However, they were collected in a similar study on a routine involving comparative analogies, and positive student performance effects were found related to a classroom teacher's use of the routine (Bulgren et al., 2000). Additionally, this study did not focus on the effects of the routine on students' ability to make comparisons independently.

The results of this investigation and other studies related to other Content Enhancement Routines (e.g., Bulgren et al., 1988, 2000) suggest that using these instructional routines can help teachers meet the challenges associated with teaching students complex information and higher order thinking skills in inclusive content classrooms. Because this research was conducted in secondary general education classes in which diverse groups of learners were enrolled, such research can make a contribution to improvement efforts in secondary schools. Therefore, the research has implications for pre-service teacher training and ongoing staff development efforts designed to help teachers and students respond to the challenges in today's schools to meet standards and become literate citizens.

References

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Bean, T. W., Singer, H., Sorter, J., & Frazee, C. (1986). The effect of metacognitive instruction in outlining and graphic organizer construction on students' comprehension in a tenth-grade world history class. *Journal of Reading Behavior, 18,* 153–169.

Beissner, K. L., Jonassen, D. H., & Grabowski, B. L. (1993, January). Using and selecting graphic techniques to acquire structural knowledge. In *Proceedings of selected research and development presentations at the convention of the Association for Educational Communications and Technology,* New Orleans, LA. (ERIC Document Reproduction Services No. 362151)

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals.* New York: McKay.

Bos, C. S., & Anders, P. L. (1990). A theory-driven interactive instructional model for text comprehension and content learning. In T. E. Scruggs & B. Y. L. Wong (Eds.), *Intervention research in learning disabilities* (pp. 166–182). New York: Springer-Verlag.

Bos, C. S., Anders, P. L., Filip, D., & Jaffe, L. E. (1989). The effects of an interactive instructional strategy for enhancing reading comprehension and content area learning for students with learning disabilities. *Journal of Learning Disabilities, 22,* 384–390.

Bulgren, J. A., Deshler, D. D., & Schumaker, J. B. (1993). *The content enhancement series: The concept mastery routine*. Lawrence, KS: Edge Enterprises.

Bulgren, J. A., Deshler, D. D., Schumaker, J. B., & Lenz, B. K. (2000). The use and effectiveness of analogical instruction in diverse secondary content classrooms. *Journal of Educational Psychology, 92,* 426–441.

Bulgren, J. A., & Lenz, B. K. (1996). Strategic instruction in the content areas. In D. D. Deshler, E. S. Ellis, & B. K. Lenz (Eds.), *Teaching adolescents with learning disabilities: Strategies and methods* (2nd ed., pp. 409–473). Denver, CO: Love.

Bulgren, J. A., Lenz, B. K., Deshler, D. D., & Schumaker, J. B. (1995). *The Concept Comparison Routine.* Lawrence, KS: Edge Enterprises.

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1988). Effectiveness of a concept teaching routine in enhancing the performance of LD students in secondary-level mainstream classes. *Learning Disability Quarterly, 11,* 3–17.

Clarke, J. H. (1991). Using visual organizers to focus on thinking. *Journal of Reading, 34,* 526–534.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Deshler, D. D., & Schumaker, J. B. (1988). An instructional model for teaching students how to learn. In J. L Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 391–411). Washington, DC: National Association of School Psychologists.

Dickson, S. V., Simmons, D., & Kameenui, E. J. (1995). Instructional design: Instruction in expository text. A focus on compare/contrast structure. *Learning Disability Forum, 20*(2), 8–15.

Doyle, C. S. (1999). *The use of graphic organizers to improve comprehension of learning disabled students in social studies.* Unpublished master's thesis, Kean University, Union, NJ.

Dunston, P. J. (1992). A critique of graphic organizer research. *Reading Research and Instruction, 31,* 57–65.

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis, 11,* 189–196.

Horton, S. V., & Lovitt, T. C. (1989). Using study guides with three classifications of secondary students. *Journal of Special Education, 22,* 447–462.

Horton, S. V., Lovitt, T. C., & Bergerud, D. (1990). The effectiveness of graphic organizers for three classifications of secondary students in content area classes. *Journal of Learning Disabilities, 23,* 12–22.

Irvin, J. L. (1990). *Vocabulary knowledge: Guidelines for instruction. What research says to the teacher*. Washington, DC: National Education Association.

Joint Committee on Teacher Planning for Students With Disabilities. (1995). *Planning for academic diversity in America's classrooms: Windows on reality, research, change, and practice.* Lawrence: University of Kansas, Center for Research on Learning.

Lenz, B. K., Bulgren, J. A., & Hudson, P. (1990). Content enhancement: A model for promoting the acquisition of content by individuals with learning disabilities. In T. E. Scruggs & B. L. Y. Wong (Eds.), *Intervention research in learning disabilities* (pp. 122–165). New York: Springer-Verlag.

Lenz, B. K., Schumaker, J. B., & Deshler, D. D. (1991, March). *Planning in the face of academic diversity: Whose questions should we be answering?* Paper presented at the American Educational Research Association Conference, Chicago.

Mayer, R. E. (1987). *Education psychology: A cognitive approach.* Boston: Little, Brown.

McDonnell, L. M., McLaughlin, M. J., & Morrison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academy Press.

Moore, D. W., & Readence, J. E. (1984). A quantitative and qualitative review of graphic organizer research. *Journal of Educational Research, 78,* 11–17.

National Council for the Social Studies. (1990). *Social studies curriculum-planning resources.* Dubuque, IA: Kendall/Hunt.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics* (chap. 2). Retrieved October 29, 2001, from http://standards.nctm.org/documents/index.htm

Nolet, V. W., & McLaughlin, M. J. (2000). *Accessing the general curriculum: Including students with disabilities in standards-based reform.* Thousand Oaks, CA: Corwin Press.

Putnam, M. L., Deshler, D. D., & Schumaker, J. B. (1992). The investigation of setting demands: A missing link in learning strategy instruction. In L. J. Meltzer (Ed.), *Strategy assessment and instruction for students with learning disabilities: From theory to practice* (pp. 325–351). Austin, TX: Pro-Ed.

Schumaker, J. B., Deshler, D. D., & McKnight, P. C. (1991). Teaching routines for content areas at the secondary level. In G. Stover, M. R. Shinn, & H. M. Walker (Eds.), *Interventions for achievement and behavior problems* (pp. 473–494). Washington, DC: National Association of School Psychologists.

Swanson, H. L. (1990). Intelligence and learning disabilities: An introduction. In H. L Swanson & B. Keogh (Eds.), *Learning disabilities: Theoretical and research issues* (pp. 23–40). Hillsdale, NJ: Erlbaum.

U.S. Department of Education. (1997). *Nineteenth annual report to Congress on implementation of the Individuals With Disabilities Education Act.* Washington, DC: Author.

Wechsler, D. (1974). *Wechsler Intelligence Scales for Children—Revised.* New York: Psychological Corporation.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson Psychoeducational Battery.* Allen, TX: DLM Teaching Resources.

Received August 16, 2000
Revision received October 31, 2001
Accepted November 1, 2001 ■