Pergamon

Available online at www.sciencedirect.com
SCIENCE DIRECT

Learning and Individual Differences 14 (2004) 91–105

Learning and Individual Differences

www.elsevier.com/locate/lindif

# Focusing attention to deep structure in math problems: Effects on elementary education students with and without attentional deficits

Suneeta Kercood[a,*], Sydney S. Zentall[a], David L. Lee[b]

[a] *Department of Educational Studies, Purdue University, West Lafayette, IN, USA*
[b] *Department of Educational and School Psychology and Special Education, Pennsylvania State University, University Park, PA, USA*

Received 8 April 2003; received in revised form 2 September 2003; accepted 10 September 2003

## Abstract

The purpose of this study was to examine the effects of actively categorizing math problems on the math problem solving performance of students with and without attention deficit disorder (ADD). To this purpose, 52 fifth and sixth grade students were involved in actively categorizing or noncategorizing (control) tasks followed by two problem solving activities. The results of this study indicated that students more frequently organize math problems by categories based on deeper structures (i.e., concepts or operations) when they were informed of the features to look for than students who organized math problems on their own. However, those students who actively organized the math problems and formed categories on their own had higher accuracy in an assessment of generality in a subsequent problem solving task than students who were earlier given categories. These effects were similar for both groups and provided generality to prior research on recall performance to assessing problem solving.


*Keywords:* Attention; Math; Students; Attention deficit

## 1. Introduction

Active categorization interventions have been used to improve flexibility, recall, and problem solving performance of average learners. Students have been asked to categorize problems based on how they

* Corresponding author. College of Education, Butler University, 4600 Sunset Avenue, Indianapolis, IN 46208, USA. Tel.: +1-317-940-8429.

*E-mail address:* skercood@butler.edu (S. Kercood).

1041-6080/$ - see front matter 
doi:10.1016/j.lindif.2003.09.001

would solve them (Smith, 1988), identify problems within a set that were different from the others (Low & Over, 1989), or organize and categorize problems that could be solved similarly to model problems (Hardiman, Dufrensne, & Mestre, 1989). In each of these studies, the initial ability to categorize problems correlated significantly with recall performance.

Similarly in memory tasks, students with attention deficit hyperactivity disorder (ADHD), learning disabilities (LD), or reading disabilities (RD) showed improved recall of words when using active categorization. For example, Torgesen (1977) demonstrated that asking elementary grade students to sort words and pictures produced greater gains in reading recall for students with LD who were poor readers than for students without LD. Similarly, when 10–16-year-old students with RD and ADHD were asked to read words aloud and then place words with similar meanings (or that were alike in some way) into the same box, their recall performance after the above sorting activity was equivalent to that of comparisons (August, 1987). Although they were not asked to verbally state their categorizing scheme, they were provided with one possible way to categorize words (i.e., similar meanings). Unfortunately, the findings in this study could be attributed to practice with the task. That is, neither of the above memory studies included a comparison nonsort (NS) condition; thus, differential gains could simply be attributed to additional exposure to the words to be recalled.

As well, there is some indication that mental age may play a role in whether students need to have categories provided to them in memory tasks (Becker & Morrison, 1978). That is, although students had higher recall of words when asked to self-generate categories *and* when categories were provided to them compared with a noncategorization control activity, students with lower mental ages (i.e., students with LD whose mental ages were between 5 and 9 years) performed better when they were provided categories than when they were required to generate categories. Only those students with higher mental ages (i.e., between 9 and 11 years) benefited from being allowed to generate categories.

This memory performance has been improved with categorization, and students with lower mental ages made more gains when the categories were provided for them than when they generated their own. One reason children with lower mental ages may have difficulty generating categories may be because they focus on the surface features, such as names of objects, tasks, and people. This could be due to lack of experience looking beyond the surface features or it could represent a perceptual bias. For example, poor mathematical problem solvers focused on descriptor terms in the problems and then looked for the actual equations that could be manipulated to yield an answer (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Mestre, Dufresne, Gerace, & Hardiman, 1988; Mestre & Gerace, 1986). In contrast, good problem solvers or "experts" identified the "deep structure" (i.e., the underlying concepts in the math problems, e.g., problems based on Newton's laws or kinetic energy) as an aid to solve those problems. Individuals who focus on the deeper structure of problems tend to have better conceptual knowledge and greater procedural skill (Hiebert & Wearne, 1986) to apply to novel problems (Krutestskii, 1976). Specifically, in mathematics, poor problem solvers were given mathematical training in categorizing problems like the experts (i.e., with categories provided). Under these conditions, they made more correct judgements regarding solution and made fewer errors in problem solving (Hardiman et al., 1989; Schoenfeld & Herrmann, 1982).

In a related nonmathematical problem solving study, elementary grade students with and without hyperactivity were given pictures with three matrices (letters, flowers, and numbers), each of which varied along several dimensions (e.g., letters could be differentiated on letter size, type, position, and shape of the background; Tant & Douglas, 1982). Students were required to ask the examiner questions regarding the matrices, and students with hyperactivity, more than comparisons, asked a larger proportion of questions about single dimensions of a category (e.g., size only). Thus, these findings could similarly indicate that students with hyperactivity or ADHD would have greater difficulty generating their own categories.

An attentional preference for salient or novel features of stimuli (for recent review, see Zentall, 1993) may lead (a) to greater difficulty focusing on relevant stimuli that are neutral, subtle, small/detailed, or embedded within tasks and (b) to shorter perceptual cycles being established in daily interactions with the environment. Thus, students with ADHD may not attend to relevant stimuli and fail to build conceptual knowledge needed for some problem solving tasks (Tant & Douglas, 1982). In support, prior research documented that students with ADHD had specific difficulties with math concepts (e.g., time and distance) even when group differences in both Math Computation ability and IQ were statistically controlled (Zentall, Smith, Lee, & Weiczorek, 1994).

In summary, lower mental ages, less experience, and perceptual bias may all result in difficulty generating categories. Interventions of active categorization have improved problem solving performance of average learners; however, for students with mild disabilities, the available research has failed to control for repeated exposure to the materials. To advance the literature in this area, this study assessed the effects of a categorizing activity on math problem solving performance while providing a comparison NS condition that provided equivalent exposure to the task materials. We predicted that students with attentional disorders and hyperactivity would benefit more from an active categorizing activity in their math problem solving performance than would comparison students.

## 2. Method

### 2.1. Participants

#### 2.1.1. Nomination procedures

Administrators, such as school principals, vice principals, and school counselors, from eight schools in three Midwestern school districts were asked to nominate students from fifth and sixth grades who (a) had already been diagnosed as having ADD/ADHD without concurrent LD or (b) had six or more of the characteristics of inattention and impulsivity as stated in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994; A checklist of the DSM-IV characteristics was provided to the school administrators). Those same administrators were also asked to randomly select students who were average functioning and had not previously been identified with ADD, ADHD, or LD. Seventy-two of the nominated participants returned the parent and student permission letters. (Issues of confidentiality precluded collecting information on the number of students who were initially nominated in each group.)

#### 2.1.2. Identification procedures

A review of the school-based identification procedures indicated that students with ADHD were identified/labeled by a physician who used either (a) parent and teacher ratings as well as an assessment of achievement or (b) parent ratings only. Both these above procedures were sufficient to provide the students with access to services under Section 504 (Office of Civil Rights, 1988). Although our participants with ADHD had previously been identified by the physician based on one of the above criteria and 10 were currently taking medication, we assessed the current behavioral status of all the students nominated for the ADD/ADHD group using ratings on the Conners Rating Scales-Revised (CRS-R; Goyette, Conners, & Ulrich, 1978), which were obtained from a parent (Conners Parent Rating Scale-R, CPRS-R) and the child's teacher (Conners Teachers Rating Scale, CTRS-R). Because students in sixth grade did not have a class teacher, ratings for sixth grade students were obtained from the homeroom teacher. The attention deficit group (Group A) was composed of students who received a *t* score of 65 or above (i.e., 1.5 S.D. above the mean) on the Hyperactivity Index from at least one rater (i.e., the parent or the teacher) on either the CTRS or the CPRS.

From the students with permission letters who met criterion of at least average intellectual ability [i.e., >85 on the Cognitive Skills Index (CSI) from the Indiana Statewide Testing for Educational Progress (ISTEP), a nationally norm-referenced test; CTB/McGraw-Hill, 1987], the selection procedures yielded a student sample that included 29 students in Group A (27 Caucasian and 2 African American students; 20 male, 9 female). Eight of the students from Group A met criterion on both parent and teacher ratings (CTRS from 65 to 91 and CPRS from 67 to 94). Three students from Group A were on psychostimulant medication. Thirteen additional students met criterion on teacher ratings only with CTRS *t* scores ranging from 65 to 98 (interpreted as "much above average" to "very much above average" on the Hyperactivity Index). Their corresponding parent ratings were CPRS scores of 42–62 (corresponding to "average" to "slightly above average" ratings on the scale). One student from Group A was on stimulant medication. Eight remaining students previously identified with Group A met criterion based on parent ratings only, with CPRS scores of 65–98 ("much" to "very much above average" on the Hyperactivity Index). Their corresponding CTRS teacher ratings were from 40 to 55 (indicative of average ratings on the Hyperactivity Index). It should be noted that six of these eight students were on psychostimulant medication, which could explain their current failure to also meet criterion scoring on the teacher rating scale.

The comparison group (Group C) included 23 students who did not have a preexisting diagnosis of LD or ADD/ADHD and had a *t* score 50 or below on both CTRS and CPRS (22 Caucasian and 1 Hispanic American students; 10 male, 13 female). The students included in Group C had *t* scores of 40–50 on the CTRS ($M=45.04$, S.D. = 3.33) and 35–50 on the CPRS ($M=45.44$, S.D. = 6.95) Hyperactivity Indices. These *t* scores can be interpreted as in the "below average" to "slightly below average" range on the Hyperactivity Index.

Analyses of the demographic data indicated that there were no differences in gender between the groups [$\chi^2(1,\ n=52)=3.40,\ P<.10$]. Significant between-group differences were yielded on ISTEP scores on the Reading Vocabulary, Language Mechanics, Math Computation, Math Application, and CSI (an index of IQ; see Table 1). Students were also administered four subtests of the KeyMath-Revised during the first testing session to assess their current level of functioning in arithmetic and problem solving (Connolly, 1988). Students in Group A performed worse than those in Group C on the Numeration, Addition, and Problem Solving subtests (see Table 1).

### 2.2. Experimental setting, tasks, and conditions

Two experimental sessions were conducted during the child's regular school day in a testing room in the school (3.4 × 5.2 m). In this room, the student was seated at a right angle (90°) from the examiner and faced a wall with no pictures or posters.

At the start of Session 1, each student was individually administered four subtests of the KeyMath-Revised (Numeration, Addition, Multiplication, and Problem Solving), which took about 15–20 min. If a student needed a break at this time, they were permitted to walk around, get water, use the bathroom, etc.

Table 1
Group differences in demographic data

| | *df* | *F* | Group C ($n=23$) | | Group A ($n=29$) |
|---|---|---|---|---|---|
| | | | Group means (S.D.) | | |
| Age (months) | 1,50 | 0.33 | 142.84 (8.24) | < | 145.27 (9.22) |
| CTRS index | 1,50 | 41.25** | 45.04 (3.33) | < | 67.20 (16.25) |
| CPRS index | 1,50 | 38.88** | 45.44 (6.95) | < | 69.52 (15.04) |
| ISTEP/CTBS | | | | | |
| Reading Vocabulary | 1,47 | 4.24* | 107.04 (10.29) | > | 100.17 (12.39) |
| Reading Comprehension | 1,47 | 1.21 | 104.00 (10.40) | > | 100.50 (11.47) |
| Language Expression | 1,47 | 2.02 | 104.71 (12.93) | > | 99.53 (12.39) |
| Language Mechanics | 1,46 | 14.05** | 112.25 (10.63) | > | 94.71 (18.84) |
| Math Computation | 1,47 | 3.84* | 107.38 (10.90) | > | 98.80 (17.63) |
| Math Application | 1,46 | 7.48** | 108.61 (11.20) | > | 99.30 (12.01) |
| CSI | 1,46 | 9.42** | 108.20 (11.94) | > | 96.28 (14.11) |
| KeyMath-Revised | | | | | |
| Numeration | 1,50 | 8.32** | 112.17 (6.54) | > | 104.48 (11.30) |
| Addition | 1,50 | 3.79* | 107.17 (9.50) | > | 99.31 (17.40) |
| Multiplication | 1,50 | 2.03 | 103.91 (5.63) | > | 98.62 (17.05) |
| Problem Solving | 1,50 | 5.64* | 106.30 (6.60) | > | 98.44 (14.70) |

CSI from the Comprehensive Tests of Basic Skills (CTBS) and ISTEP test scores.

* $P < .05$.

** $P \leq .01$.

Each student was then given a set of six cards (7.5 × 17.5 cm). Each card contained one arithmetic (addition or multiplication) word problem that involved concepts of money, distance, or time (e.g., Lisa traveled for 2 months. If every month the travel was 8 days, how many days did Lisa travel? Tina earned 4 dimes. If Tina earned another 7 dimes, How many dimes did Tina earn? Mike biked 5 laps. If there are 7 blocks per lap, how many blocks did Mike bike?). These problems were computer generated and were based on the guidelines specified by Zentall et al. (1994). The word problems had an average of 17.3 words (range 16–19) and a mean grade level of 2.2 (assessed by the Flesch-Kincaid reading grade level, available in Microsoft Word version 6.0). Students were asked to read these problems aloud and complete the activities for one of three experimental conditions. Each condition used the same set of problems for each of two trials and ∼ 5–10 min to complete each trial.

### 2.2.1. Free Sort (FS) condition

In the FS condition, students were asked to sort problems on their own, allowing them to organize the problems based on the characteristics or structures they identified. Students were given the following instructions: "I want you to read the following problems aloud. When you are done, place these problems into piles so that when it is time for you to solve them, it will be easier for you." After a student placed the problems in piles, the examiner pointed to each of the piles and asked the student, "Why did you place these cards in this pile?" The identical procedure was repeated for a second trial.

### 2.2.2. Directed sort (DS) condition

In the DS condition, the examiner had preidentified the characteristics and asked the students to organize the math problems by the examiner guidelines. In the first trial, students in the DS condition were asked to sort the problems conceptually (i.e., by distance, time, and money). In the second trial, the students were asked to sort the math problems by similarities in operations (i.e., multiplication). Students were given the following instructions, "Put all the problems that involve multiplication here" (the examiner placed a note card with the word multiplication on the table) and "place all the problems that do not involve multiplication here" (the examiner pointed to a blank spot on the table about 25 cm away from the first note card). After each trial, the examiner pointed to each of the piles and asked the student to give a rationale for placing problems in a particular pile.

#### 2.2.3. NS condition

In both trials in the NS condition, students were asked to perform a matching-to-sample task. A paper containing the same set of problems used by the students in the other two conditions was placed on the table in front of the participant. The problems were typed single spaced in a single column with double spacing between problems. The participants were also given a set of problem cards that were the same as the problems printed on the paper. The following instructions were given, "I want you to read the problems aloud. When you are done, place the problems on these cards with the same problems found on this paper." The same procedure was repeated for the second trial.

### 2.3. Design

Students had been randomly assigned to the three conditions with the constraint that an equivalent number of students from each population and gender group was represented in each condition. Equivalent representation of students was documented by nonsignificant Group × Condition interactions for reading scores: Reading Vocabulary [$F(5,46) = 0.46$, $P = .6336$], Reading Comprehension [$F(5,46) = 0.31$, $P = .7329$], Total Reading [$F(5,46) = 0.38$, $P = .6860$], and CSI [$F(5,46) = 1.04$, $P = .3621$] and math scores: Math Computation [$F(5,46) = 2.96$, $P = .062$], Math Application [$F(5,46) = 0.54$, $P = .5877$], and Total Math [$F(5,46) = 2.77$, $P = .740$].

### 2.4. Postexperimental tasks

#### 2.4.1. Post-test problems

Immediately after completing the activities for an experimental condition in Session 1, the students were given a test set of math word problems. These test problems were a new set of six problems developed in a manner similar to the experimental problems, with an average of 17.5 number of words (range 16–19) and a mean grade level of 2. The student was asked to read the problems aloud from typed index cards (7.5 × 17.5 cm) and solve them on paper.

#### 2.4.2. Wisconsin Card Sorting Test (WCST)

A week later in Session 2, students were individually administered the WCST (Heaton, Chelune, Talley, Kay, & Curtis, 1981). The WCST provides a measure of sorting and problem solving activity. In addition to good reliability (range $r = .91-.96$; Axelrod & Henry, 1992), this test has been shown to differentiate between students with ADHD and typical students in problem solving, learning, visual spatial, and categorization abilities (Boucugnani & Jones, 1989; Chelune, Ferguson, Koon, &

Dickey, 1986; Pineda et al., 1998). The WCST was given after the experiment so that it would not confound the effects of the conditions and would provide an assessment of generality of the conditions.

The WCST consists of 4 stimulus cards and 128 test cards containing sets of geometric designs that vary in color, form, and number. The participant is presented with the four stimulus cards and asked to sort the remaining deck of test cards, matching them to one of these four stimulus cards. The sorting principle of color, form, or number is predetermined by the test and is changed by the examiner without informing the participant when s/he achieves 10 correct responses. The participant was not informed of which sorting principle was in effect but was provided with standardized accuracy feedback of "right" or "wrong" after each card placement.

### *2.5. Measures*

#### *2.5.1. Accuracy*

Sorting accuracy for Session 1 was defined as the total number of word problem cards sorted correctly (i.e., by concepts or operation) for each of the two trials of the two sorting conditions (FS and DS). For the NS task, accuracy was recorded as the total number of cards matched correctly. Problem solving accuracy was defined as the total number of math problems solved correctly.

#### *2.5.2. Time*

Duration (s) taken to complete the word problem sorting tasks and the NS (card matching) task was recorded for each trial.

#### *2.5.3. Sorting categorical responses*

Students' verbal responses that provided a rationale for their choice of word problem sorting were taped during Session 1 and later ranked using a five-level scoring system developed by the authors. Level 0 was defined by sorting descriptions that contained irrelevant explanations (such as "I am not sure why" or "I just think it should be placed here"). Level 1 was identified by use of superficial words in the verbal responses (such as "how many" or "how far"). Level 2 was sorting descriptions consisting of concept-related words such as travel, hours, dollars, cents, and miles. Level 3 was defined by sorting descriptions that directly used mathematical concept words such as time, money, and distance. Level 4 was defined by responses referring to the operations of multiplication or addition.

#### *2.5.4. WCST performance*

The measures obtained during Session 2 from the WCST were based on the test scores in each of the following areas: (a) trials to criterion, (b) total number of errors, perseverative responses, and errors (i.e., repeating responses related to color, shape, or number), (c) number of categories (e.g., color and form/shape) completed or identified correctly out of six possible (i.e., 10 cards must be identified correctly within a category before moving to a new category), (d) failure to stay within a category (i.e., failure to consistently identify the sorting principle of color, shape, or number), (e) nonperseverative responses and errors (i.e., other errors), and (f) conceptual level responses (correct sorting principal). These measures have been used in the past to demonstrate problem solving, abstract thinking, attention, and categorization abilities (e.g., Pineda et al., 1998).

### 2.5.5. Behavior

Arm movements were measured by recordings on an Actometer (Timex motion recorder 108, a modified self-winding calendar wristwatch) that was strapped around a student's nondominant wrist. This device is sensitive to any large movements of the nondominant arm (e.g., arm raised in the air, reaching out or up, or moving the desk) but not the hand or fingers alone. Actometers have been used to provide a reliable and valid measure of student activity (see Madan-Swain & Zentall, 1990; Zentall, 1989).

### 2.5.6. Inter-rater reliability

A graduate student, who was naive to group status and to experimental conditions, was trained in behavioral observation using video recordings of the same behaviors from prior research studies. Once there was a 90% agreement between the graduate student and the examiner on the video behavioral recordings, the graduate student randomly selected 20% of the student sample to observe and record concurrently with the examiner. Interobserver agreement was calculated by the ratio of the total number of agreements to the total number of observations. The frequency of bottom torso movements had a mean of 76.8% agreement between raters (range from 66% to 100%). A graduate student also rescored 20% of student protocols of the math problem solving task and of taped student responses to the questions, with inter-rater agreement for accuracy of problem solving at 100% and categorical responses at 96.6% (range from 92% to 100%). The protocols for the WCST were computer scored and inter-rater reliability for these scores was not recalculated. Agreement was also not obtained for the time measures.

## 3. Results

Equivalence in the cells of the design was found for IQ [$F(5,43) = 1.04$, $P$=.362]; however, there was a trend for pretest differences in Math Computation ability [$F(5,43) = 2.96$, $P$=.064]. For this reason, accuracy, time, and problem solving data were analyzed using analyses of covariance (ANCOVA), with Math Computation as a covariate. Before analyzing the data, homogeneity of regression slopes for all the measures was documented through nonsignificant Group × Covariate interactions. Nonsignificant results reported without exact probabilities are $F < 2$. Normal probability plots were used to assess univariate normality. Although (a) visual inspection of the normal probability plots revealed departures from normality for both behavioral and performance measures and (b) log and square root transformations failed to correct for that skewness, ANCOVA was used because it is robust against departures from normality (Glass & Hopkins, 1984).

Two planned contrasts were conducted for each condition effect[1]: (1) a comparison of sorting (i.e., FS and DS) versus NS conditions and (2) a contrast between the two sorting conditions (i.e., FS vs. DS). As indicated by Keppel (1982, p. 146), the number of planned comparisons was restricted to 1 less than the number of conditions and evaluated at the standard level of significance ($P < .05$). Simple effects of each trial were also examined only where there were identical procedures across trials (i.e., the FS condition).

[1] Keppel (p. 106) has indicated that "...analytical comparisons can be conducted on a set of data without reference to the significance or nonsignificance of the omnibus $F$ test."

### 3.1. During experimental conditions

#### 3.1.1. Sorting performance

*3.1.1.1. Accuracy.* Because all students in the NS task matched the problems correctly, data from this condition were not included in the accuracy analysis. Thus, the data for only the two sorting conditions were subjected to a 2 (Group: A and C) × 2 (Condition: FS and DS) × 2 (Trial) ANCOVA, with Math Computation scores the covariate. Main effects were not found for Group ($F<2$) or Condition [$F(1,25)=3.10$, $P<.10$]. Simple effects ANCOVA of condition showed that in Trial 1 students in the DS condition were more accurate ($M=5.41$, S.E. = 0.44) than students in the FS condition [$M=3.33$, S.E. = 0.54; $F(1,25)=8.33$, $P<.01$, $d=1.18$].[2] However, this difference was not significant by Trial 2.

*3.1.1.2. Time.* The time data were examined with a 2 (Group: A and C) × 3 (Condition: FS, DS, and NS) mixed-design ANCOVA. (Trial was not included in this overall analysis due to the noncomparability of procedures between the two trials within each condition.) Only in the FS condition were the two trials comparable. For this analysis only, we found that the students performed similarly between the two trials.

In the overall analysis across trials, planned contrasts showed that students in the FS condition took longer ($M=215.59$, S.E. = 17.81) to complete the tasks than students in the DS condition [$M=154.34$, S.E. = 14.69; $F(1,42)=6.85$, $P<.01$, $d=0.86$], although there was no time difference between the sorting (FS and DS) and the NS conditions. Simple effects analysis of Trial 1 also indicated that students in the FS condition took longer ($M=119.17$, S.E. = 10.79) than students in the DS condition ($M=72.14$, S.E. = 8.90) to complete the tasks [$F(1,42)=11.00$, $P<.01$, $d=1.18$]. By Trial 2, students in both sorting conditions took longer ($M=89.3$, S.E. = 8.05) than the students in the NS condition [$M=67.8$, S.E. = 6.79; $F(1,42)=5.87$, $P<.05$, $d=0.73$]. However, by Trial 2, there were no time differences between FS and DS.

*3.1.1.3. Sorting categorical responses.* Because of infrequent responses to the categorical levels, Levels 0–2 were combined as a low-level response (i.e., those based on surface features of a problem) and Levels 3 and 4 were combined to be high-level responses (i.e., those based on deeper features of a problem such as concepts/operations). A direct logistic regression analysis with odds ratio (Tabachnick & Fidell, 1996, pp. 608–610) was performed on the outcome probability of scoring a Level 2 (or high level) response with three predictors: groups, conditions, and trials. Logistic regression is essentially a multiple regression analysis but with a dichotomous dependent variable.

Results indicated that the probability of scoring high-level responses (i.e., on the bases of concepts or operation) was higher for those students in the DS condition ($M=0.8$, S.D. = 0.06) than for those in the FS condition [$M=0.6$, S.D. = 0.15; $\chi^2(1, n=52)=6.17$, $P$=.01, $d=1.82$]. There were no group differences or changes in student responses over trials [$\chi^2(1, n=52)<2$].

[2] Standardized mean differences were calculated using pooled S.D.s of the groups being compared (Olejnik & Algina, 2000). Cohen (1988) suggested the following terms for evaluating values of $d$: small effect = 0.2, medium effect = 0.5, and large effect = 0.8.

## 3.2. Postexperimental tasks

### 3.2.1. Problem solving

Accuracy and time data were similarly analyzed by a 2 (Group) × 3 (Condition) ANCOVA but without the factor of trial because there was only one trial.

*3.2.1.1. Accuracy.* Analysis of the accuracy data yielded no differences among conditions [$F(1,42) = 2.44$, $P < .10$]. Planned contrasts showed that students who were in the FS condition solved more problems correctly ($M = 5.4$, S.E. = 0.39) than students in the DS condition ($M = 4.5$, S.E. = 0.32); however, this effect failed to achieve statistical significance [$F(1,42) = 3.08$, $P < .10$, $d = 0.85$]. No differences were found between the combined sorting (FS + DS) and the NS conditions.

Table 2
ANOVA for WCST performance

| | *df* | *F* | FS | DS | NS |
|---|---|---|---|---|---|
| | | | Group means (S.D.) | | |
| Categories completed | | | | | |
| Group | 1,46 | 0.22 | | | |
| Condition | 2,46 | 2.04 | | | |
| Group × Condition | 2,46 | 3.47* | Group C: 6.00 (0.01) | 5.50 (1.06) | 4.33 (1.65) |
| | | | Group A: 5.33 (0.70) | 4.66 (1.41) | 5.36 (1.20) |
| Total number of errors | | | | | |
| Group | 1,46 | 1.59 | | | |
| Condition | 2,46 | 1.86 | | | |
| Group × Condition | 2,46 | 4.32* | Group C: 13.50 (7.14) | 33.87 (20.91) | 40.22 (19.12) |
| | | | Group A: 37.44 (12.22) | 41.44 (23.67) | 27.90 (16.72) |
| Perseverative errors | | | | | |
| Group | 1,46 | 1.01 | | | |
| Condition | 2,46 | 1.85 | | | |
| Group × Condition | 2,46 | 2.96* | Group C: 7.16 (3.81) | 17.50 (12.30) | 20.11 (7.04) |
| | | | Group A: 17.88 (6.75) | 20.77 (14.12) | 14.45 (9.75) |
| Nonperseverative errors | | | | | |
| Group | 1,46 | 1.40 | | | |
| Condition | 2,46 | 1.06 | | | |
| Group × Condition | 2,46 | 3.57* | Group C: 6.33 (4.13) | 16.37 (11.69) | 20.11 (15.81) |
| | | | Group A: 19.55 (8.04) | 20.66 (12.42) | 13.45 (7.91) |
| Trials to criterion | | | | | |
| Group | 1,46 | 7.82** | Group C: 65.00 (6.72) | | |
| | | | Group A: 71.62 (10.29) | | |
| Condition | 2,46 | 1.57 | | | |
| Group × Condition | 2,46 | 2.26 | | | |
| Percentage of conceptual level responses | | | | | |
| Group | 1,46 | 0.85 | | | |
| Condition | 2,46 | 2.05 | | | |
| Group × Condition | 2,46 | 4.43* | Group C: 80.16 (8.94) | 63.32 (17.17) | 53.87 (17.93) |
| | | | Group A: 60.71(10.71) | 57.26 (19.92) | 66.87 (16.36) |

* $P < .10$.

** $P < .01$.

*3.2.1.2. Time.* Planned contrasts indicated that students in the sorting conditions (FS + DS) took longer ($M$ = 238.8, S.E. = 23.27) to solve the problems than students in the NS condition [$M$ = 156.41, S.E. = 19.50; $F(1,42) = 10.93$, $P < .01$, $d = 1.02$]. Students in the FS condition also took somewhat but not significantly longer ($M$ = 267.53, S.E. = 25.52) than students in the DS condition ($M$ = 210.13, S.E. = 21.04) to solve the problems [$F(1,42) = 2.93$, $P < .10$, $d = 0.58$].

*3.2.1.3. WCST.* The data from the WCST were analyzed using a 2 (Group) × 3 (Condition) analysis of variance (ANOVA). Only the statistically significant effects and not the trends are discussed below.

*3.2.1.4. Trials to criterion.* This refers to the measure Total Number Correct on the WCST and has been relabeled in this article to indicate what this measure actually assessed. Students in Group A required more trials ($M$ = 71.62, S.D. = 10.29) to achieve criterion than Group C [$M$ = 65.0, S.D. = 6.72; $F(1,46) = 7.82$, $P < .01$, $d = 0.74$]. The comparisons who had been in the FS condition during experimental conditions took fewer trials ($M$ = 64.83, S.D. = 4.16) to achieve criterion on the WCST than students in Group A [$M$ = 78.11, S.D. = 8.83; $F(1,15) = 12.43$, $P < .01$, $d = 1.80$].

*3.2.1.5. Categories completed.* Simple effects analyses of the 2 (Group) × 3 (Condition) interaction (see Table 2) indicated that comparisons, who were assigned to the sorting conditions, completed more categories ($M$ = 5.75, S.D. = 0.53) than those comparisons assigned to the NS condition [$M$ = 4.33, S.D. = 1.65; $F(1,20) = 7.27$, $P = .01$, $d = 1.29$]. Similarly, students from Group C who were earlier in the FS condition completed a higher number of categories postexperimentally ($M$ = 6.0, S.D. = 0.01) than students in Group A [$M$ = 5.33, S.D. = 0.70; $F(1,15) = 5.20$, $P < .05$, $d = 1.22$]. Differences were not significant for Group A, nor were there differences between the sorting conditions (i.e., FS vs. DS) for either group or between groups in any of these conditions (see Fig. 1).

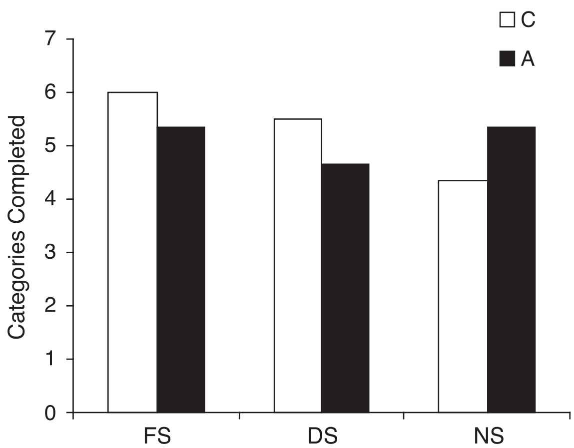


Fig. 1. Total number of categories completed by Groups A and C on the WCST. These students had earlier participated in the FS, DS, or NS experimental condition.

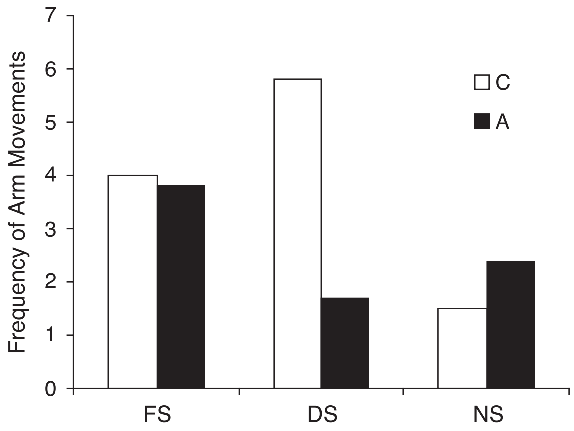


Fig. 2. Frequency of arm movements of students in Groups A and C during the postexperimental math problem solving task. These students had earlier participated in the FS, DS, or NS experimental condition.

### 3.3. Behavior

The behavioral data, which were assessed postexperimentally during the math problem solving task, were analyzed by a 2 (Group) × 3 (Condition) ANOVA. This analysis yielded a Group × Condition interaction for arm movements [$F(2,46)=3.13$, $P<.05$; see Fig. 2]. Simple effects for each group showed that comparisons had higher arm movements in the sorting conditions (FS + DS; $M=4.9$, S.D. = 3.9) than the NS condition [$M=1.5$, S.D. = 2.1; $F(1,20)=4.91$, $P<.05$, $d=1.02$]. The comparisons also had more arm movements ($M=5.8$, S.D. = 4.9) in the DS condition than students in Group A [$M=1.7$, S.D. = 2.3; $F(1,14)=4.60$, $P=.05$, $d=1.09$]. For students in Group A, there were no differences between sorting and NS, nor were there differences between the two types of sorting conditions for either group.

## 4. Discussion

Active categorization/sorting interventions have been used to improve the flexibility, organization, and recall of material as well as problem solving performance of average learners. For students with LD, RD, and ADHD, findings have indicated improved recall of words from the use of sorting without controlling for additional exposure to the task stimuli. Furthermore, categorization used with problem solving had not previously been examined. Because students with attentional deficits have difficulty with specific math concepts and with forming categories, we proposed that they would benefit more from this sorting/categorizing activity than would comparison students. That is, the sorting activity would focus the attention of the students to the relevant dimensions of the problems. If differential findings for the students with ADHD were yielded, they could *not* be attributed to active responding because active responding was included in each of the three conditions.

During the experimental conditions, all students in the DS condition in the first trial were better able to correctly identify the concept or operation of the math problem than students in the FS condition. This finding supports prior research assessing recall performance for students with LD with mental ages of 5–

9 years, who performed better with provided categories than with self-generated categories (Becker & Morrison, 1978). As well, this finding provides some generality to our problem solving task for students with and without attentional disorders from prior research on memory performance with students with LD, RD, and ADHD (Becker & Morrison, 1978; Torgesen, 1977).

However, the improved problem solving performance in the DS condition was not significant by the second trial. This may have been due to the fact that during the second trial of the DS condition, students were required to switch their schema of categorization (i.e., changing from categorizing problems by concepts such as money, distance, and time in the first trial to operations such as multiplication and addition in the second trial). Thus, when the schema of categorization was changed during the second trial of DS, the continued advantage of providing categories was lost. In addition, while a schema for organizing and categorizing (DS) increased the accuracy of all students' scores in categorizing material in the immediate setting and during the first trial of the experimental sessions, it did not increase the students' problem solving performance after experimental sessions (i.e., failed to generalize).

An analysis of the total time taken to complete the tasks during the experimental session showed that all the students took longer to complete the FS than the DS task on the first trial (i.e., before the required shift in the DS condition). Furthermore, in the FS condition, where change was not required, students maintained their system of categorization from Trial 1 to Trial 2, similar to the findings of Chi et al. (1981) who concluded that FS patterns are not ad hoc but are based on prior learning. After the experimental session, students who had been assigned to the FS condition also took longer in a subsequent math problem solving assessment than those students previously assigned to the DS condition. In fact, students in both sorting conditions (FS and DS) took longer to complete their math problems than students who had previously experienced the NS condition. These findings indicate some carry-over of processing time from the prior sorting experimental conditions to this postexperimental assessment.

The WCST, a measure of the categorization/sorting and problem solving ability, was administered to the students ~ 1 week after conducting the experimental and postexperimental tasks as an assessment of generality. In prior research, this measure had been sensitive to differences between students with and without ADHD specifically on the subtests of trials to criterion, number of categories identified, errors, and conceptual level responses (e.g., Boucugnani & Jones, 1989; Pineda et al., 1998). Replicating past research, this sample of students with attentional deficits required more trials to criterion (i.e., referring to Total Number Correct on the WCST) than did the comparisons.

The Group × Condition interactions further indicated that students with attentional deficits performed differently from comparisons in response to their prior exposure to the experimental conditions. Contrary to predictions, however, these differences were attributable to the comparison students whose performance was differentially improved by prior sorting experience. Specifically, only those comparison students, who had previously experienced the sorting conditions, completed more categories on the WCST.

In summary, the results of this study demonstrated that students, who were given advanced notice of particular features of math problems, identified those features more easily and faster than students who were asked to generate the features of math problems on their own. However, these gains were specific to the experimental session and did not generalize to a subsequent problem solving task. When students were required to form their own categories, they took longer on the subsequent problem solving task than students who were earlier provided with a schema of categorization by the examiner. This increase in processing time was offset by a trend for improved problem solving accuracy (supported by effect sizes in the moderate to high range). The improvement for all the students on this assessment of immediate generality could be attributed to their deeper cognitive processing on initial trials of the experimental math task or to a carry-over of method. In an assessment of generalization to a different type of task, significant differential benefits from the prior week's experience with sorting versus NS conditions were only to be found for typical students. Overall, educational implications are that strategies involving grouping or sorting are useful for typical children but may be selectively more difficult to learn for students with ADD/ADHD.

There are several possible explanations for why students with attentional problems failed to show the predicted gains that were clearly found for their classmates in this assessment of generalization. Because students with attentional deficits have more difficulty forming categories than their peers, they may have needed more sorting time than 5–10 min, more trials, or more direct instruction or directed questioning on the salient features of problems/tasks. Future research could be conducted extending instructional time into short practice sessions (i.e., to avoid problems with sustaining attention) to see if the gains we reported for typical children could also be demonstrated for the ADHD group. Difficulty with sorting tasks could be explained by the requirements of delayed responding while sorting (Zentall, 1993).

Advances of this study over prior work include those related to design (i.e., the control of active responding and practice in each condition) and the addition of tasks to assess generalization. Statistical control included the partialling out the effects of Math Computation ability while assessing mathematical problem solving for students with attention disorders and for their classmates.

## Acknowledgements

This research was conducted as a part of the dissertation requirements of the first author, and the results of this study were presented at the Annual Meeting of the American Educational Research Association in New Orleans, LA, April 2000.

We greatly appreciate the feedback provided by Mary Campbell, Department of Psychological Sciences, Purdue University. We thank the administrators, counselors, teachers, parents, and children from the Lafayette, Tippecanoe, West Lafayette, Rossville, and Delphi school corporations, who made this study possible.

## References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Axelrod, B. N., & Henry, R. R. (1992). Age-related performance on the Wisconsin Card Sorting Test, Similarities and Controlled Oral Word Association Tests. *Clinical Neuropsychologist*, *6*, 16–26.

August, G. J. (1987). Production deficiencies in free recall: A comparison of hyperactive, learning disabled and normal children. *Journal of Abnormal Child Psychology*, *15*, 429–440.

Becker, L., & Morrison, G. (1978). *The effects of levels of organization on clustering and recall in normal, learning disabled and educable mentally retarded children*. Final Report under U.S. Office of Education, Bureau of Education for the Handicapped Grant G007604709, Eric Document Reproduction Service ED 180 147.

Boucugnani, L. L., & Jones, R. W. (1989). Behaviors analogous to frontal lobe dysfunction in children with attention deficit hyperactivity disorder. *Archives of Clinical Neuropsychology*, *4*, 161–173.

Chelune, G. J., Ferguson, W., Koon, R., & Dickey, T. O. (1986). Frontal lobe disinhibition in attention deficit disorder. *Child Psychiatry and Human Development*, *16*, 221–234.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Cohen, J. (1988). *Statistical power analysis of the behavioral sciences* (2nd ed.). New York: Academic Press.

Connolly, A. (1988). *KeyMath-Revised*. Circle Pines, MN: American Guidance Service.

CTB/McGraw-Hill (1987). Monterey, CA: CTB/McGraw-Hill.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normal date on revised Conners Parents and Teachers Rating Scales. *Journal of Abnormal Child Psychology*, *6*, 221–236.

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory and Cognition*, *17*, 627–638.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1981). *Wisconsin card sorting test manual*. Odessa, FL, USA: Psychological Assessment Resources.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199–223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Keppel, G. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

Krutetskii, V. A. (1976). *The psychology of mathematical abilities in school children*. Chicago: University of Chicago Press.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

Low, R., & Over, R. (1989). Detection of missing and irrelevant information within algebraic story problems. *British Journal of Educational Psychology*, *59*, 296–305.

Madan-Swain, A., & Zentall, S. S. (1990). Behavioral comparisons of liked and disliked hyperactive children in play contexts and the behavioral accommodations by their classmates. *Journal of Consulting and Clinical Psychology*, *58*, 197–209.

Mestre, J. P., Dufrensne, R., Gerace, W., & Hardiman, P. T. (1988). *Hierarchical problem solving as a means of promoting expertise*. Washington, DC: National Science Foundation (Eric Document Reproduction Service ED 310 931).

Mestre, J. P., & Gerace, W. J. (1986). *Studying the problem solving behavior or experts and novices in physics via computer-based problem-analysis environments*. Eighth Annual Conference of the Cognitive Science Society (pp. 741–746). Hillsdale, NJ: Lawrence Erlbaum Associates.

Office of Civil Rights, W. J. (1988). *Free appropriate public education for students with handicaps: Requirements under Section 504 of the Rehabilitation Act of 1973*. Washington, DC: Office of Civil Rights.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241–286.

Pineda, D., Ardila, A., Rosselli, M., Cadavid, C., Mancheno, S., & Mejia, S. (1998). Executive dysfunctions in children with attention deficit hyperactivity disorder. *International Journal of Neuroscience*, *96*, 177–196.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 484–494.

Smith, M. U. (1988). *Expertise, mental representation and problem solving success: A study of categorizations of classical genetics problems by biology faculty, genetic counselors and students*, Eric Document Reproduction Service ED 292 674.

Tant, J. L., & Douglas, V. I. (1982). Problem solving in hyperactive, normal and reading disabled boys. *Journal of Abnormal Child Psychology*, *10*, 285–306.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: HarperCollins College.

Torgesen, J. K. (1977). Memory processes in reading disabled children. *Journal of Learning Disabilities*, *18*, 350–357.

Zentall, S. S. (1989). Attentional cueing in spelling tasks for hyperactive and comparison regular classroom children. *Journal of Special Education*, *23*, 83–93.

Zentall, S. S. (1993). Research on the educational implications of attention deficit hyperactivity disorder. *Exceptional Children*, *60*, 143–153.

Zentall, S. S., Smith, Y. N., Lee, Y. B., & Weiczorek, C. (1994). Mathematical outcomes of attention deficit hyperactivity disorder. *Journal of Learning Disabilities*, *27*, 510–519.