COGNITION AND INSTRUCTION, *15*(4), 435–484

Complex Mathematical Problem Solving by Individuals and Dyads

Nancy J. Vye and Susan R. Goldman
Learning Technology Center
Vanderbilt University

James F. Voss and Cindy Hmelo
Learning Research and Development Center
University of Pittsburgh

Susan Williams
Learning Technology Center
Vanderbilt University

Cognition and Technology Group at Vanderbilt
Vanderbilt University

Complex mathematical problem solving was examined in 2 studies using an episode from *The Adventures of Jasper Woodbury.* Each episode in the *Jasper* series consists of a narrative story that ends with a complex challenge that students are to solve. Solving the challenge involves formulating subproblems, organizing these subproblems into solution plans, differentiating solution-relevant from solution-irrelevant data, coordinating relevant data with appropriate subproblems, executing computations, and deciding among alternative solutions. The episode examined in these studies was *The Big Splash.* The challenge is to construct a business plan for a booth at a school fun-fair fund-raiser. This article reports the results of using a technique that we developed for analyzing complex problem solving: solution-space analysis. In Experiment 1, the performances of 6th-grade and college students solving the problem under think-aloud instructions are compared. Relative to 6th-grade students, college students were more likely

Requests for reprints should be sent to Nancy J. Vye, Box 45–GPC, Vanderbilt University, Nashville, TN 37203. E-mail: vyenj@ctrvax.vanderbilt.edu

to generate solution attempts and correct solutions and to consider multiple-solution plans. Both groups of students were highly accurate in generating important subgoals. They were equally unlikely to evaluate time and money constraints involved in the solution. In Experiment 2, dyads of 5th graders solved the same problem as in Experiment 1, with instructions to work together to reach a solution. The solution-space analysis was augmented by a focus on the argumentation processes manifest in the problem solving of the dyads. Among the dyads, more successful problem solving was associated with more coherent argument structures in the problem-solving dialogues. Coherence was reflected in (a) goals giving rise to attempts, (b) attempts giving rise to new goals, and (c) goal-appropriate calculations. In addition, many of the dyads in Experiment 2 explored multiple-solution paths. Discussion focuses on characteristics of problems that make solutions difficult, the kinds of reasoning that dyadic interactions support, and considerations of instructional environments that would facilitate the kinds of problem-solving and reasoning processes associated with coherent solutions.

In recent years, policy makers, as well as educational researchers, have emphasized the need to foster instructional programs that enhance the development of thinking and problem-solving skills within specific subject domains (American Association for the Advancement of Science, 1989; Cognition and Technology Group at Vanderbilt [CTGV], 1990; Glaser, 1994; Resnick, 1989; Voss & Means, 1991). Such trends in educational reform reflect a shift in theoretical viewpoints away from the view that learning is the accumulation of facts and associations. Increasingly, educators have come to view learning as a constructive activity in which knowledge is interpreted, structured, and adapted to new situations—with the assumption that all learning revolves around thinking (Resnick, 1989).

Consistent with this view of learning, research on problem solving and expertise has demonstrated that the development of expertise is not so much a function of the application of general heuristic strategies (Newell & Simon, 1972) as it is the acquisition of principled, structured knowledge (Chi, Glaser, & Farr, 1991) coupled with the concomitant development of self-regulatory and monitoring skills (Glaser, 1991). In contrast to novices, experts characteristically develop rich problem representations, reason more effectively within their domain, and are more facile at acquiring related knowledge (Glaser, 1994; Resnick, 1989; Voss, 1991; Voss, Blais, Means, Greene, & Ahwesh, 1986). These findings suggest that, if we want students to move successfully from novice to higher levels of performance, we need to create instructional environments wherein facts, algorithms, or concepts do not lie inert (Whitehead, 1929) but rather become integrated, applied, and put to effective use.

Faced with the challenge of constructing meaningful thinking environments that promote understanding, researchers have increasingly begun to reemphasize the importance of viewing knowledge as a tool to be used in meaningful ways (Dewey, 1933) and of viewing learning as being essentially a social activity (Vygotsky, 1934/1986). Such conceptualizations have prompted researchers to devise classroom-centered studies in which students are provided with meaningful activities—that is, authentic tasks and materials in which students are encouraged to work in a collaborative or an apprenticeship fashion (Barron et al., 1995; Brown, Collins, & Duguid, 1989; CTGV, 1990, 1991, 1992, 1993, 1994, 1997; Lamon et al., 1996; Palincsar & Brown, 1984; Scardamalia & Bereiter, 1991; Secules, Cottom, Bray, Miller, & The Schools for Thought Collaborative, 1997).

In the area of mathematics research in particular, several classroom-centered studies have been designed in which students are engaged in collaborative problem solving in an effort to develop mathematical understanding (Lampert, 1986; Mason, Burton, & Stacey, 1982; Schoenfeld, 1985, 1987). In these studies, students actively engage in problem-solving activities in which problem solutions are not presented but rather proposed, defended, and negotiated. In the work of Lampert, for example, efforts to promote understanding of the mathematical principles underlying multiplication are centered in instructional environments in which the teacher and the students engage in purposeful activities. The students are encouraged to construct meaningful relations between quantities, to invent numerical procedures, to generate numerical representations, and, all the while, to provide rationales for the hypotheses they generate.

Establishing learning environments such as Lampert's (1986) is a long-term process that depends on a teacher who is astute about both pedagogy and mathematics (Heaton & Lampert, 1993). Recent assessments indicate that many teachers are ill-prepared to create these kinds of environments (e.g., Ball & Rundquist, 1993; Cobb, Wood, Yackel, & McNeal, 1992; Peterson, Carpenter, & Fennema, 1989; Peterson, Fennema, & Carpenter, 1991; Peterson, Fennema, Carpenter, & Loef, 1989; T. Wood, Cobb, & Yackel, 1992) and that it often takes several years of observation and reflection for teachers to effect fundamental changes in their own pedagogy (Ball & Rundquist, 1993; Heaton & Lampert, 1993). Part of the problem in creating environments that promote mathematical understanding is having materials that afford the kinds of discussions and challenging problems that engage students. Another equally important problem is having the knowledge of mathematics necessary to guide discussions in classrooms such as Lampert's. The approach that we have taken to the materials issue is to create video-based materials that pose complex situations and problems for students to solve (CTGV, 1990, 1991, 1992, 1993, 1994, 1997). *The Adventures of Jasper Woodbury* poses complex problems that involve mathematical problem solving. When these videos are used in the classroom, they enable certain kinds of activities but by no means assure that the activities will take place (CTGV, 1991, 1993, 1994; Goldman & CTGV, 1991; Van Haneghan et al., 1992). (Although we do not discuss it here, we are currently using these materials to create contexts for deepening teachers' understanding of the mathematics involved in the episodes. Initial efforts in this direction are discussed in Barron et al., 1995.)

Studying the development of generative knowledge and problem solving in complex problems poses a set of interesting challenges. First, there are few conceptualizations of complex problem solving beyond the enumeration of general processes and activities (e.g., see Fay & Klahr, 1996). Second, ways must be devised to measure the underlying processes that take place as children move through these complex problem spaces. Thus, as Greeno (1986) pointed out in response to Lampert's (1986) work, we need to develop more detailed analyses of the process and content knowledge that students acquire, which will enable success in mathematics learning. Previous results indicate that there is a considerable difference between successful performance on well-structured math problems, such as two-digit calculations, and successful performance on complex math problems, such as those presented in the *Jasper* series (CTGV, 1992, 1993, 1994, 1997; Goldman & CTGV, 1991; Van Haneghan et al., 1992).

Solutions to *Jasper* problems involve multiple goals that have a hierarchical structure, numerous constraints, multiple-solution options, and multiple-solution paths. Some of the cognitive processes involved in solving *Jasper* problems include formulating the subproblems needed to solve the overall problem, organizing the subproblems into solution plans, coordinating relevant data with appropriate subproblems, distinguishing relevant from irrelevant data, formulating computational procedures to solve subproblems and the overall problem, and determining the feasibility of alternative plans. Traditional school environments produce students who are ill-prepared to solve problems requiring the coordinated use of such processes; presumably because of this, *Jasper* problems are difficult to solve. For example, when college students individually attempted to solve two of the *Jasper* trip-planning problems, they dealt adequately with only about 60% of the solution space, often failing to optimize their solution by considering multiple alternatives (CTGV, 1993, 1994; Goldman & CTGV, 1991). Middle school age students who score extremely well on standardized mathematics achievement tests (90th percentile and above) also have difficulties similar to those of college students. In addition, middle school age students experience difficulties (a) determining the data that are relevant to solving particular subproblems and (b) successfully formulating computational procedures for such subproblems.

Because individuals have so much difficulty solving these problems, they provide an interesting context for collaborative problem solving. Triads of 11- and 12-year-olds solving the same trip-planning problem included more of the elements of the problem-solving space in their solutions than did the previously mentioned college students and high-achieving sixth-grade students (Barron, 1991). Although Barron's research provided no analyses of the processes enabling collaborative groups to perform better than individuals, she speculated that the interactions among the members of the triad would predict individual differences among the groups. Several studies support this speculation (Goldman, Cosden, & Hine, 1992;

Hine, Goldman, & Cosden, 1990; Webb, 1989; Webb, Ender, & Lewis, 1986). In the mathematics area, Webb (1989, 1991) indicated that, when dyad members can provide meaningful explanations to their partners, problem-solving performance improves over individual performance.

A second possibility with respect to why individuals solving the trip-planning problems failed to consider multiple plans—and thereby a higher proportion of the problem solution space—concerns the semantics of the trip-planning problems. Solvers may have felt that, once they had a solution, they had met the requirement. A different problem domain might have made more salient issues of optimization and the need to evaluate alternative plans. Specifically, business plans often present multiple options and comparative analyses. By their very nature, business plans involve optimization. That is, alternative plans are often feasible but carry different levels of cost and risk. Frequently, these factors must be weighed against one another, as in cost–benefit comparisons.

The research reported in this article describes two studies of complex problem solving in the context of one *Jasper* business-planning adventure, *The Big Splash*. The primary purpose of Experiment 1 was to examine whether individuals solving complex problems in this domain experienced difficulties similar to those experienced when solving trip-planning problems. Optimization was of particular interest, including feasibility and alternative plan consideration because these issues are more explicitly part of business planning than of trip planning. In Experiment 1, individuals produced think-aloud protocols while solving *The Big Splash*. We interpret performance in the individual situation as a baseline or benchmark that can be used to index performance by other students or under different task circumstances. As in our earlier studies (CTGV, 1994, 1997; Goldman & CTGV, 1991), the verbalizations tended to describe the goals and data students were considering but provided relatively limited information about the reasoning and decision-making processes that they were using.

In Experiment 2, we had dyads cooperatively solve *The Big Splash*. We stressed with students the need to talk to one another in order to cooperate on a solution. We expected that the students' interactions would produce information that could be used to chart their explorations of the solution space in a more complete way than was possible with the individuals' verbalizations. Our goal in Experiment 2 was to understand the nature of the argumentation and reasoning processes associated with different solution outcomes. Due to pragmatic constraints on the students' available time, we focused students on the part of the problem concerned with expenses and profit. (In Experiment 1, students focused on revenue, expenses, and profit.) Although we recognize that these task differences preclude our making definite claims about the effects of working in dyads relative to individuals, in our discussion, we highlight interesting differences in the patterns of findings for dyads and individuals and speculate as to underlying mechanisms that could account for these differences.

For purposes of examining the extent of the solution space that was considered by students in each of the studies, we applied the method we had devised in our earlier studies of trip planning: The problem was analyzed as a solution space consisting of the major elements or goals that needed to be considered, their interrelations, and decision or evaluation points (CTGV, 1993, 1994). (This analysis is an adaptation of planning net analysis as described by VanLehn & Brown, 1980.) Based on the content of the think-aloud protocols and any written work, students' problem solving can be characterized in terms of the elements of the solution space considered. In this context, this analysis was applied to students participating in both experiments.

In addition, in Experiment 2, the interactions between the members of the dyads supplied the information regarding components of the solution process. We also looked at the *process* of solving. In so doing, we assumed three major processes: establishing goals, reasoning that helps specify the steps needed to satisfy such goals, and the application of appropriate mathematical skills. We refer to this framework as the *GRA model,* and we characterize dyads' solution processes in terms of the goals stated, the reasoning as indexed by arguments used during the solution process, and the application of appropriate mathematical computations. We considered the contingency relations among goals, arguments, and computations—and how these related to elements of the problem space—to address questions of coherence and goal-directedness of the solution process. We were also concerned with how the dyad interaction was reflected in the problem-solving process. To examine this, we focused on the contribution of each member of a dyad, as well as on the dyad interaction, to identify conditions that facilitate the solution process.

EXPERIMENT 1: COLLEGE AND HIGH-ACHIEVING SIXTH-GRADE MATH STUDENTS' BUSINESS PLAN SOLUTIONS

College and sixth-grade high-achieving math students were shown the *Jasper* adventure, *The Big Splash.* They worked by themselves to solve the problem, providing think-aloud protocols while they worked.

Method

Participants

Participants were 14 high-achieving sixth-grade students and 16 college undergraduates. The sixth graders ($M = 11$ years, 9 months) were recruited from a local public school. All had scored at or above the 80th percentile on the mathematics portion of the Tennessee Comprehensive Achievement Program (administered approximately 6 months prior to the study), with 8 scoring at the 99th percentile. The college undergraduates were volunteers from an introductory psychology course at a southeastern university where the average Scholastic Assessment Test score is above 1,050. They received their choice of course credit or monetary remuneration for participating.

Materials

The Big Splash is an approximately 15-min video about a young teen named Chris who wants to set up a dunking booth at his school fair. Figure 1 contains a summary of the story. Briefly, Chris's idea is to sell tickets for the opportunity to dunk teachers in a pool of water. The money earned from the booth would help purchase a video camera for the school's TV studio. The principal of Chris's school, Ms. Stieger, offers to lend money to students who wish to have booths at the fair. The maximum amount of each loan is $150, and to qualify, the estimated income from the booth must be double the estimated expenses. To be eligible for the loan, interested students must prepare a business plan detailing their estimated income and expenses and their compliance with these rules (see Frame 6 in Figure 1).

In the video, Chris gathers the information that he will use to develop his plan. He conducts a survey of students at his school to determine if they would be interested in buying a ticket to dunk a teacher and, if so, how much they would be willing to pay. Chris also explores where to get the dunking machine and pool and various options for filling and emptying the pool. Chris gathers information relevant to several means of filling the pool, including using the school hose, having the fire department deliver the water, and buying water from the pool store. Similarly, there is information in the video relevant to several methods of emptying the pool (e.g., the pool store can remove it, the city's public works truck can be used, or the water can be siphoned or drained out).

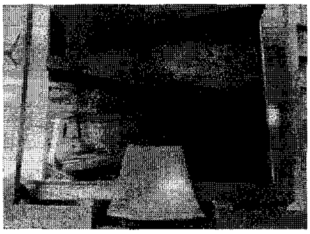
The problem posed at the end of the video is to prepare the business plan that Chris should present to Ms. Stieger. All information needed to solve the challenge has been embedded in the story; however, as students watch the video for the first time, they do not know what problem they will have to solve. Accordingly, it is difficult for the student to determine what data are going to be relevant to the challenge. Data are presented as a natural part of the story. For example, in *The Big Splash,* Chris is shown taking a survey and summarizing the data. This information is relevant to estimating revenue for the dunking booth, and after students find out what the challenge is, they will need to realize the relevance of these data and to use them appropriately (see the text accompanying Frames 8 and 9 in Figure 1).

The Big Splash Story Summary

Chris walks to the local fire station to do some research for a report he is writing. As he walks toward the fire hall, he sees a city truck cleaning the streets with water. The truck has the following sign on it:

> *Department of Public Works*
> Capacity 3,000 gal

When Chris arrives at the fire hall, he meets Chief Sullivan, the fire chief. Chief Sullivan shows Chris the pumper truck, which can pump 1000 gallons a minute, assuming they are hooked up to a hydrant. Chris asks if they usually find a hydrant, and Chief Sullivan says that usually they do, assuming they are in the city. Chief Sullivan says that the pumper truck holds *some* water (the sign says "capacity 700 gallons"), but it doesn't hold enough water to do much fire fighting. Chief Sullivan then shows Chris the 38-foot ladders. Chris asks how often the firefighters go out on fires. Chief Sullivan says that they average 20 to 30 calls a week, although some of the calls are false alarms.

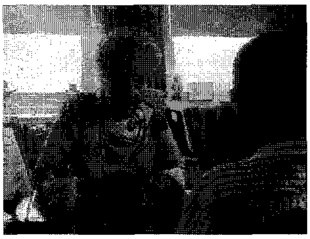


Chris sees a "weird looking contraption" in the fire hall. Chief Sullivan says that it is a dunking machine that they built. He says that they sometimes sell 100 tickets per hour. They rent it out for $25.00 per day, and the proceeds go to their scholarship fund. Chris asks where they put the water. Chief Sullivan says that they set it up in an above-ground swimming pool. Chris then asks the chief to see the pole which firefighters slide down when there is a fire.

FIGURE 1 (Continued)

Thursday afternoon, as Chris sits in school, he listens to an announcement about the upcoming Fun Fair. The principal, Ms. Stieger, announces that the fair will be held a week from Friday from 10:00 A.M. to 3:00 P.M. on the athletic field. She says that the proceeds will go toward a new video camera for the student television station. She also says that they hope to raise $800, and they need at least one more good money-making project for the fair. The school will lend someone up to $150.00 to cover the initial costs of the project. Plans need to be given to her by Wednesday.

As Chris listens to the announcement, he daydreams about his teacher getting drenched as he is dunked by a dunking machine. Two formulas, $V = \pi r^2 \times h$ and $\pi \approx 3.14$, are shown on the board behind the teacher.



Later that afternoon, Chris goes to Ms. Stieger to find out what he needs to do to have a booth in the Fun Fair. She tells him that he needs to write a business plan that describes how much money he expects to take in, his gross revenue, and how he arrived at that number. Secondly, she needs an itemized account of all his expenses. The expenses cannot exceed $150.00, the maximum amount she can lend him. The plan needs to include everything he will need so that she can see that everything will be in the right place at the right time. Her rule of thumb is that, if everything looks good and the revenue is at least twice his expenses, the project is viable and she can make him the loan.

Later, Chris meets Jasper at the Soda Shop to talk about his plan. Jasper asks Chris how many students go to his school. Chris has already researched this and found that there are 380 students enrolled in the school and, on an average day, twenty are absent. Jasper says it would be nice to know how many students would buy tickets. Chris estimates that more than half probably would. Chris and Jasper decide a survey would be a good way of getting a more accurate estimate.

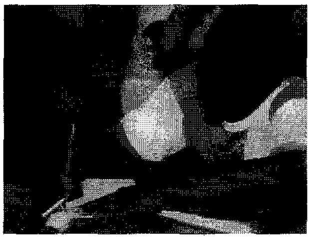
FIGURE 1 (Continued)

Chris hands out surveys to every sixth student in line going into the school cafeteria during lunchtime on Friday. The survey asks the following:

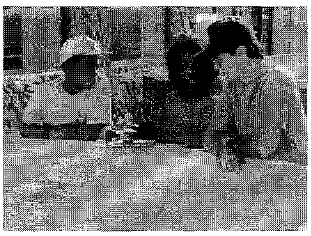
> Sometimes at fairs, you'll see one of those dunking machines. When you hit the target with a ball, a person will fall in the water.
>
> 1. Would you like to "dunk" one of your teachers at the All -School Fair?
>
>   *Circle one*
>   YES or NO
>
> 2. What's the most you would pay for one ticket (2 throws)?
>
>   *Circle one*
>   A. $ .50 C. $1.50
>   B. $1.00 D. $2.00

That afternoon, Chris meets with Jasper again at the Soda Shop. They look at the results of Chris' survey. Chris says that 58 out of 60 students would like to dunk a teacher. He shows Jasper the rest of the results:

$ .50	13 kids
$1.00	21 kids
$1.50	16 kids
$2.00	8 kids

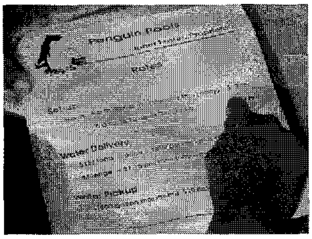


They begin to figure out how much money Chris would make if he charged the different amounts for tickets based on his survey results. Jasper says that all 58 students would pay at least $ .50 for a ticket. Chris multiplied 58 times $ .50 and got $29.00. Next, Chris says that all but 13 students would spend at least $1.00 for a ticket. To determine how many students would pay $1.00 for a ticket, Chris adds 21, 16, and 8 and gets an answer of 45. Jasper has to leave, so Chris continues to find what the best ticket price would be on his own.



On Saturday, Jasper and Chris meet Janet Foster, the proprietor of Penguin Pools. She says that she usually does not rent pools, but since Chris is doing this for a school project, she might make an exception. Ms. Foster shows them a pool that she thinks might be just what he needs. It is 3 feet deep and 12 feet in diameter. It holds about 2500 gallons of water. She will rent it to Chris for $40 a day, in advance, and she will give him one-fourth off for the second day if he needs it. Jasper asks if this price includes delivery.

FIGURE 1 (Continued)



Ms. Foster shows them a price sheet:

Set-up

Hours: 6 a.m. to 6 p.m. (Monday thru Friday) $7/hr
6 a.m. to 12 noon (Saturday) $10/hr

Water Delivery

$15/load (load = 1,500 gal) plus mileage
Mileage = $1.15/mi (one way; return trip free)

Water Pickup

(1,500 gal maximum) $10 flat fee

Ms. Foster says that Harold, her set up man, figures it takes about two hours to set a pool up or take it down, and he will start at 6 a.m., but he won't work past 6 p.m. Chris asks if two hours includes filling the pool. Ms. Foster says, "No, it doesn't. So be sure to allow yourself plenty of time up front to fill the pool, especially if you are going to fill it from a hose." Ms. Foster points out that Chris could buy the water from her. It costs $15 a load plus mileage; her truck holds 1500 gallons. She says that it takes about 15 minutes to pump water in or out of the pool. Ms. Foster says there are 7.5 gallons in a cubic foot of water. Chris and Jasper leave the pool store with 45,836.5 miles showing on the odometer.

They arrive at the school (odometer reading: 45,845.4 mi) to test how quickly water comes out of the water hose. It takes them 20 minutes to get to the school from the pool store. When they arrive at school, they walk past a fire hydrant to get to the hose. It takes 30 seconds to fill a five gallon bucket with water from the school's water hose.

Sunday afternoon, Chris goes back to the fire hall to talk to Chief Sullivan. Chief Sullivan says that Chris can use the machine for a day, and they'll bring it over Thursday afternoon. Chris asks if they can begin filling it up at 8:30 Friday morning. Chief Sullivan agrees and says that they can use the pumper truck to fill the pool, which will only take a few minutes. Chris won't have to pay anything for the fill up because it's a school project. He emphasizes that if the firefighters are out on a call, they won't be able to help until they get back.

FIGURE 1 (Continued)

Tuesday afternoon, Chris and Jasper meet at the Soda Shop. Jasper is reading the newspaper. Chris sees an article that says that the Public Works Department will drain swimming pools free of charge. Concerned citizens should call the Mayor's Office between 8 a.m. and 5 p.m. Chris says that he will be presenting his plan to Ms. Stieger the next day.

Challenge

- Five teachers have told Chris that they are willing to be dunked.
- Prepare a business plan for the dunking machine project as if you were making the presentation to Ms. Stieger.

FIGURE 1 Story board and summary for *The Big Splash.*

Procedure

Students first viewed *The Big Splash.* They were advised to pay close attention to the story because they would be asked questions afterwards about the problem posed therein. After viewing the video, students were interviewed individually in a single, problem-solving session. The maximum session duration was 1.5 hr. They were given the following instructions:

> Chris's challenge is to prepare a business plan to present to Ms. Steiger so that she will lend him money to set up a dunking booth. I want you to imagine that you are Chris and to prepare a business plan for Ms. Steiger, solving all the problems that you need to decide on the best plan.

To ensure that students understood what a business plan entailed, the experimenter replayed the scene from the video in which the principal explained the meaning of the terms *income* and *expenses* and detailed her criteria for making a loan (see Frame 6 in Figure 1).

All problem-solving sessions were audiotaped. Students were asked to think aloud as they solved the problem. Just before they began solving the business plan challenge, the think-aloud procedure was explained through the use of a word problem. Students solved the word problem, and the experimenter monitored and prompted their verbalizations as they were solving the problem.

Students were provided with paper and pencil for doing calculations, and the interviewer volunteered to perform calculations on a calculator as requested by the student. To eliminate memory demands during problem solving, students were provided with a written summary of the story, as shown in Figure 1. The summary contained all of the data (solution relevant and irrelevant) and the major story events from the video. During problem solving, the interviewer used minimal probing except to encourage thinking aloud and to elicit clarification on students' statements.

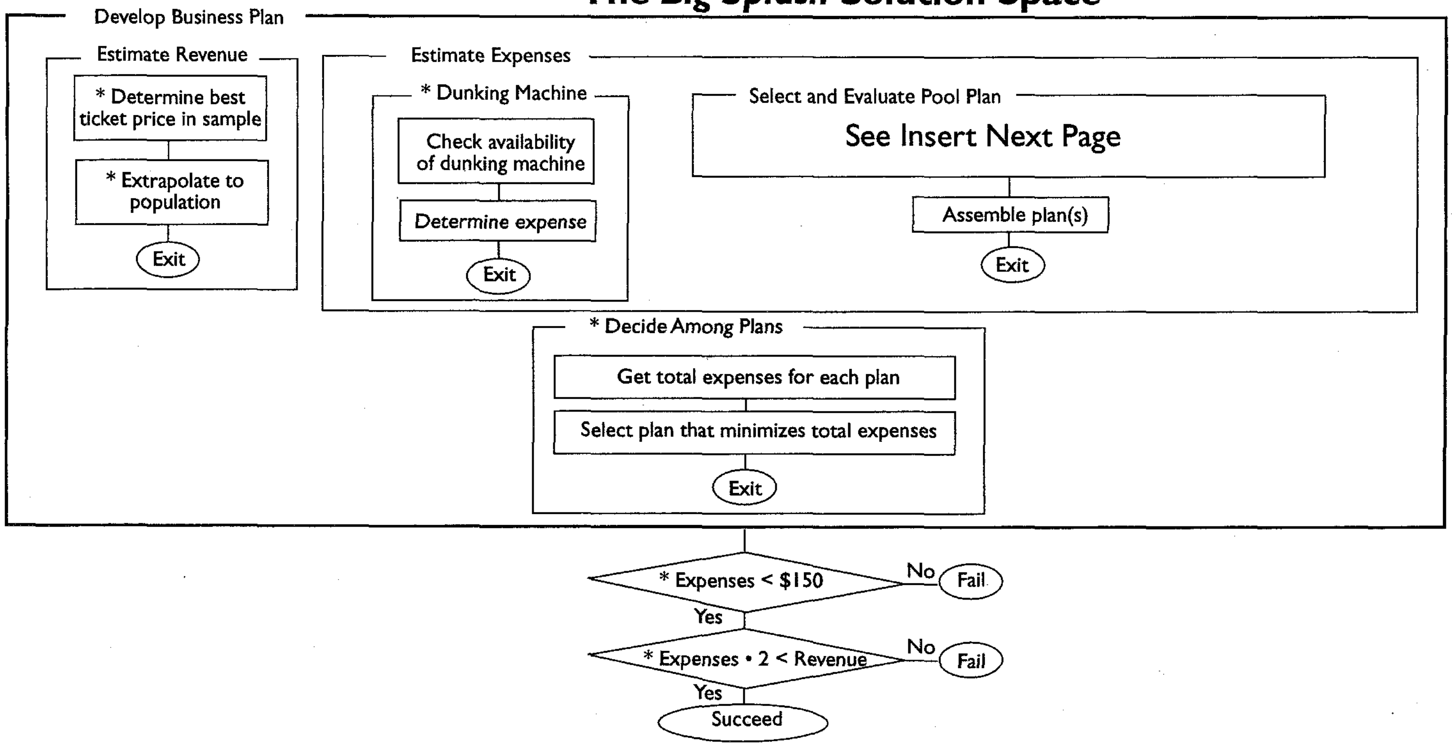
Representation of the Solution Space

Think-aloud solution protocols were examined in relation to a solution space for the problem. We captured the elements of the plan needed to solve *The Big Splash* challenge by using a planning net representation, as shown in Figure 2. (Order of execution is conveyed in the vertical arrangement of elements and not in the horizontal arrangement.) The solution involves developing a business plan by estimating revenue and expenses and then testing expenses against the two constraints set forth in the story: Expenses cannot exceed $150, and the revenue must be twice the estimated expenses. Estimating revenue involves determining the ticket price that will produce the greatest estimated income based on the sample and then extrapolating to the population to estimate total income. With respect to expenses, the best business plan involves constructing a pool plan that minimizes total costs. Pool plan costs are a function of both fixed and variable expenses. For example, the cost of the dunking machine is fixed. Different options for pool filling, however, carry different costs. Also, these options fill the pool at different rates. For some options, it takes so long that the pool must be rented for a second day, and additional rental costs are incurred. As a result, total expenses for the pool are determined by several interrelated factors.

An important aspect of the solution space is that there are multiple plans for filling the pool and for emptying the pool. Each plan needs to be tested against time, risk, and conservation constraints. If a particular plan meets these constraints, costs for the filling or emptying plan need to be determined. Optimizing the solution implies looking for as many filling or emptying plans as meet the constraints so that the overall expenses for the pool can be found. Minimizing only the filling or emptying expenses may not produce the lowest expenses for the overall pool plan because of the interdependencies among solution-space elements.

An interesting aspect of *The Big Splash* solution space is that some of the filling options have no costs but either take a long time (using the school hose) or are risky (having the fire department bring the water). There are also trade-offs for the emptying options (e.g., if the water is drained or siphoned out, it will take time and be environmentally irresponsible).

The *Big Splash* Solution Space



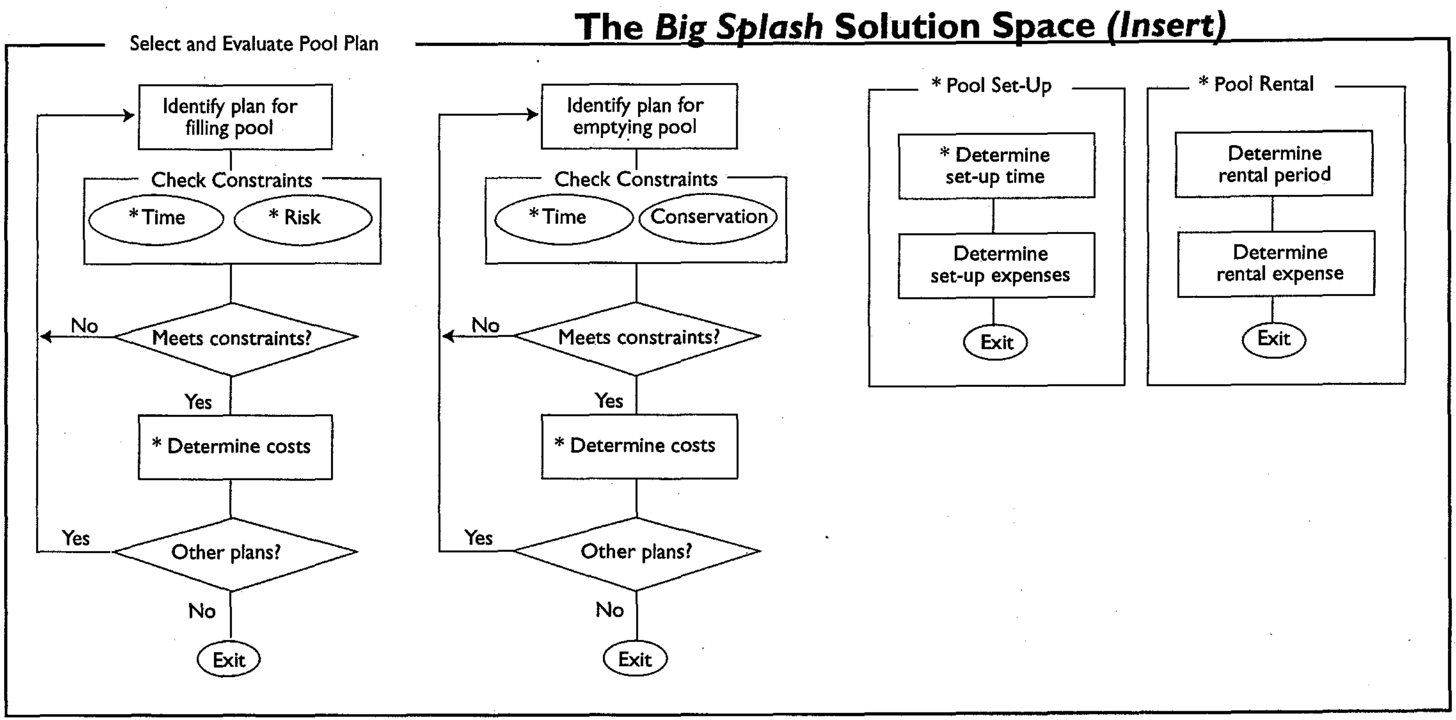


FIGURE 2 Planning net representation of the solution space for *The Big Splash*.

Results and Discussion

College students' and children's think-aloud protocols were analyzed from three perspectives on problem solving: statement types, error types, and the business plan elements reflected in Figure 2. Two coders independently scored 25% of the protocols. Coder agreement exceeded 90%. Disagreements were resolved through discussion. Scoring and results for each perspective are described separately.

Statement Types

To describe the nature of students' solution-relevant discourse during problem solving, unique protocol statements were categorized by type. (Statement repetitions were not scored.) The statement categories and their definitions were as follows:

- *Goal:* A goal describes what the student plans to do. These are often used to justify a calculation that a student is performing or deems useful to perform. It does not include plans to find solution-relevant data or givens.
- *Isolated fact:* An isolated fact is a statement about data that is not part of a calculation a student is performing or performs next.
- *Attempt:* An attempt is a procedure, for example, any arithmetic operation.
- *Solution:* A solution is the outcome of an arithmetic operation. It includes beliefs or guesses about the outcome of an operation.

The following protocol segment from one student illustrates how these criteria were applied:

> It takes 20 minutes to drive from the pool store to the school [isolated fact]. How many loads of water will he need to buy? [goal]. It holds 2,500 gallons, so that's 2 loads [attempt and solution for preceding goal].

Categorizing the protocol statements took into account the context in which the statement occurred. That is, to be credited as an attempt, the procedure or arithmetic operation had to occur in the context of a relevant goal or plan; likewise, the same had to occur for coding solution statements. Guesses or beliefs were also coded as solutions when they occurred in the context of an appropriate attempt or goal. The difference between attempt and isolated fact was the context. For example, in the preceding protocol segment, the time to drive from the pool store to the school was not part of the goal plan for determining the number of loads of water needed, and there was no preceding goal with which the statement was relevant. Hence, it was coded as an isolated fact. In contrast, if a student mentioned the time to drive from the pool store to the school in the presence of a relevant goal (e.g., the goal of figuring out how long it would take to get water from the pool store to the school), it was credited as an attempt. In other words, isolated facts were coded if the stated information was not related to the goal or plan under consideration.

Statement Type Results

The mean number of statements in each category for college and sixth-grade students is shown in Figure 3. College students provided significantly more statements than sixth graders, $F(1, 28) = 12.58$, $MSE = 26.57$, $p < .001$. There was a significant effect of statement type, $F(3, 84) = 94.12$, $MSE = 7.33$, $p < .001$. Attempt and solution statements were more frequent than were goal statements and isolated facts (Scheffé critical difference = 2.46, $p < .01$). The Statement Type × Group interaction, however, was also significant, $F(3, 84) = 11.97$, $MSE = 7.33$, $p < .001$. A Scheffé test indicated that the significant difference between groups was restricted to attempt and solution statements (critical difference = 4.40, $p < .01$).

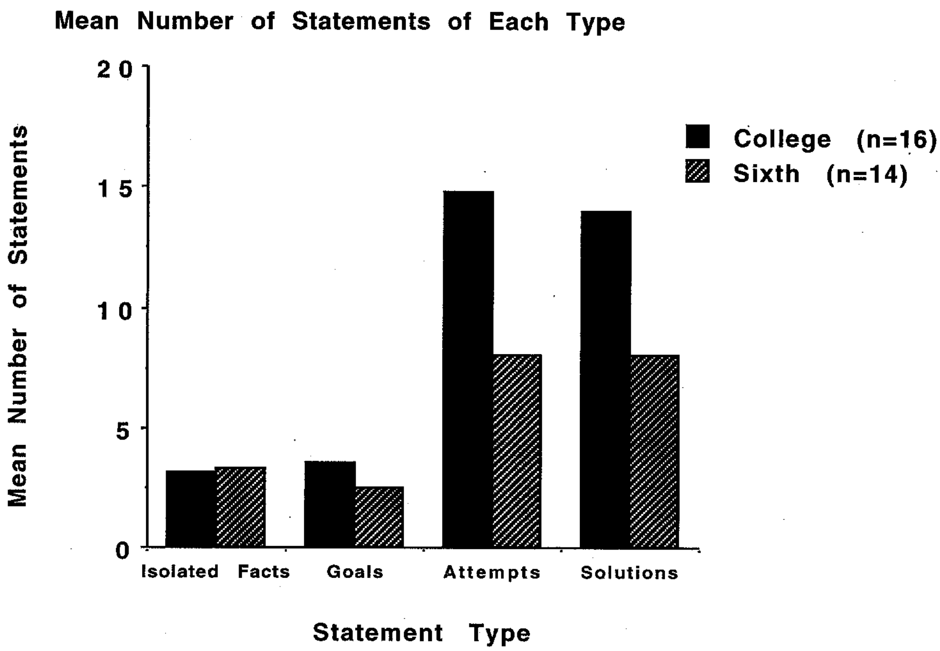


FIGURE 3 Mean number of statements per category provided in the think-aloud protocols of participants in Experiment 1.

College students' protocols contained significantly more attempt and solution statements than the sixth graders' did.

An additional analysis was conducted on just the solution statements to determine if there were between-groups differences in the number that were correct and incorrect. Correct statements were of numerical values mathematically derived from information in the video that were appropriate in the context of the problem solving. College students ($M = 10.18$) made more correct solution statements than did sixth graders ($M = 3.93$), but there were no differences in the number of incorrect solutions (Ms = 3.75 and 4.14, for college students and sixth graders, respectively); for the Group × Correctness of Solution interaction, $F(1, 28) = 13.28$, $p < .001$. Within groups, the interaction indicates that college students were three times as likely to produce correct solutions as they were to produce incorrect ones. For sixth graders, however, correct and incorrect solutions were equally likely. Finally, mathematical derivations of solutions could involve multiple steps. When we examined the number of errors per incorrect solution, we found that sixth graders produced twice as many as the college students did, $F(1, 25) = 7.21$, $p < .05$. Thus, the college students had a much higher ratio of correct to incorrect solution statements than the sixth graders did.

These data indicate that most of the information in the protocols dealt with attempts and solutions to those attempts, with relatively little discourse on what goals needed to be met. This finding may reflect the relatively high degree of redundancy that exists between attempts and goals. That is, if a problem solver is working on ticket price, it may be redundant to report "He needs to think about ticket price." It is noteworthy that the number of isolated facts was relatively low because, on occasion, we observed students who, not knowing quite where to start, begin going through all the "facts" they can think of (e.g., CTGV, 1994). In this context, such fact-listing behavior was not very common.

Types of Errors

The differences between groups in the rate of correct solutions were quite large. To determine if there were also differences in the types of errors made, we classified the errors that occurred in the solutions according to the following categories:

1. *Math formulation:* Errors that resulted from an incorrect mathematical formulation of the problem. Examples include (a) adding rather than subtracting odometer readings to determine distance and (b) adding the estimated income from each ticket price to determine the income from the population rather than doing the appropriate extrapolation procedure.

2. *Calculation:* Errors that resulted from a mistake in a computational procedure. Otherwise, the problem was correctly formulated and contained the correct data or givens.

3. *Plan element omission:* Errors that resulted from the omission of a part of the plan. Examples include failing to include (a) the cost of the dunking machine when determining total expenses or (b) the mileage charge for a second load of water from the pool store.

4. *Plan element misconception:* Errors that resulted from faulty assumptions. Otherwise, the problem was correctly formulated and contained the correct data. Examples include (a) computing an overestimate of labor charges based on the incorrect assumption that there are labor costs for filling the pool with water and (b) computing an underestimate of the cost of water from the pool store based on the incorrect assumption that partial loads of water could be purchased. These misconceptions reflect inaccurate understanding of the constraints operative in the problem space.

5. *Wrong given:* Errors that resulted from using incorrect data. Otherwise, the problem was correctly formulated. Examples include (a) using $8 instead of $7 as the hourly charge for setting up the pool or (b) using $20 instead of $25 as the cost for renting the dunk tank from the fire department.

6. *Unsubstantiated outcome:* Errors that resulted when students apparently solved a problem by guessing—that is, when there was no overt evidence that the stated information was the result of a mathematical operation. We note that it is possible the student may have performed a mental operation; hence, we call these *unsubstantiated outcomes* rather than *guessing*.

The number of errors of each type made by each subject were analyzed in a two-factor multivariate analysis of variance in which grade level (college, sixth) was the between-subjects factor and type of error was the within-subject factor (six types of errors). The main effects were significant for grade, $F(1, 28) = 12.04$, $MSE = 1.05$, $p < .01$; type of error, $F(5, 140) = 25.98$, $MSE = 0.72$, $p < .01$; and the Grade × Type interaction, $F(5, 140) = 8.26$, $MSE = 0.72$, $p < .01$. The data are provided in Figure 4. Comparisons of the means indicated that college students made fewer errors than sixth graders in two of the six error categories: plan element omission, $t(28) = 4.07$, $p < .01$, and use of wrong given information, $t(28) = 2.30$, $p < .05$. Grade differences in the other categories were not significant. The mean error rates were 2.80 and 6.07, for college students and sixth graders, respectively.

The most frequent source of errors was plan element omission among both the college (40% of the errors) and sixth-grade (55% of the errors) students. Plan element omission errors were most often associated with expenses—for example, when students forgot to include the cost of a second load of water for the pool, or the cost of disassembling the pool, or the mileage charge associated with transport-

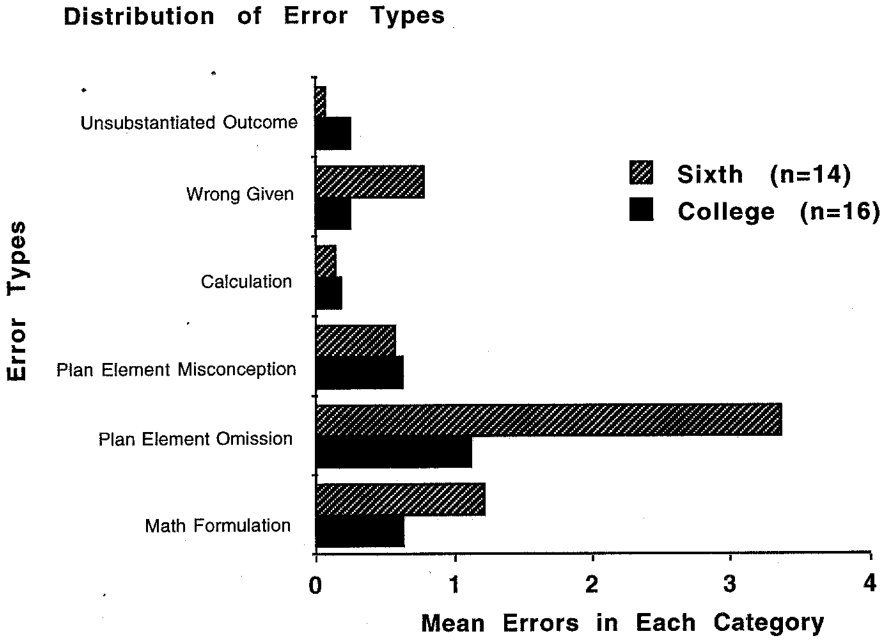


FIGURE 4 Distributions of error types in Experiment 1.

ing water for the pool. There was no single plan element that was omitted by more of the sixth graders than college students.

The second most frequent source of errors was math formulation, accounting for 22% of the college students' errors and 20% of the sixth graders' errors. Many of the formulation errors were associated with students' attempts to estimate revenue. This was the case for both college and sixth graders; however, more sixth graders made errors of this type when determining revenue. Many sixth-grade students did not seem to understand that sample data could be used to estimate population parameters. A number of students did not use the information about the proportion of people in the sample who would buy a ticket at a given price to determine the proportion of the population who would buy a ticket at this price. For example, the following is an excerpt from a sixth-grade student's protocol.

> OK. First he's got the dunking booth and then ... he has the ticket cost. So, umm ... 13 of the students say 50 cents, 21 say $1, 16 say $1.50, and 8 say $2. More people said $1 than anything else. $1 has 58 students—no, $1 times 60 students equals ... $60. And umm, $60. ... Well, Chris said that they were 380 students at his school but 20 were usually absent. That that leaves ... 380

> minus 20. That leaves 360 students that are possible that could be there. And if you multiply 360 times 1, that would be $360 that the school could make.

This student's analysis ignores the data indicating that only 45 of the 60 students polled would pay $1. Instead the student takes the highest raw frequency response and applies it directly to the school population. (Application of the correct sample proportion to the population yields an income estimate of $270.) In contrast to the sixth-grade students, all of the college students seemed to understand that they could use sample information to extrapolate to the population; their formulation errors reflected confusions regarding the appropriate ratio for representing the part–whole relation of the sample to the population.

It is interesting to note those aspects of problem solving that did not seem to be major sources of difficulty for participants in this study. Calculation errors accounted for only a small percentage of the errors made by either sixth-grade students (2%) or college students (7%). Likewise, use of inappropriate data (13% for sixth-grade and 9% for college students) and violations of the problem space constraints (8% for sixth-grade and 13% for college students) constituted relatively small percentages of the errors. The latter contrast with sources of difficulty experienced by individual solvers of the trip-planning problems (CTGV, 1994).

To summarize, problem solving tended to be incomplete rather than inaccurate. That is, mathematical procedures were generally executed accurately on the appropriate given information. The two major sources of difficulty were failures (a) to consider necessary elements of the solution space and (b) to select the appropriate mathematical procedures.

Plan Elements in the Solution Space

The final analysis examines both the plan elements comprising students' problem solving and the degree to which students optimized their plans and tested them against the constraints on expenses that had been set forth by Ms. Stieger. Twelve plan elements that reflect the major subgoals in the solution space were the focus of the analyses (see the elements preceded by an asterisk in Figure 2). For purposes of these analyses, we examined the business plan that participants said was their best. This was either their only plan or the one they said was their best when asked this question by the researchers.

Plan elements were scored with respect to three levels of inclusion:

1. *Mention:* Credit was given when the student stated or implied the need to address a specific element or when the student attempted to solve an element, even if the element had not been explicitly stated. This measure indexes awareness of the element. In previous research, we have found differences in

awareness of the importance of certain plan elements (CTGV, 1994; Goldman & CTGV, 1991).

2. *Attempt:* Credit was given when the student attempted a mathematical solution to a specific goal or subgoal. The attempt had to involve the use of mathematical operators and specific numerical values for credit to be given. No credit for an attempt was given if the student merely stated isolated facts without trying to use them in a calculation or in some other purposeful manner.

3. *Solve:* Credit was given when the student produced a correct mathematical solution. Estimated answers were coded as correct if the numerical bases for them were clear.

In reporting the results of these analyses, we have grouped the plan elements into four clusters: revenue, expenses, feasibility, and optimization.

Revenue

Estimating revenue involved estimating the amount of money that would be generated by ticket sales on the day of the fair. There were two plan elements associated with making this estimate. The first involved using the information obtained from the sample to determine a ticket price at which income would be maximized. The relevant data had been presented in the story (see Frame 9 and accompanying text in Figure 1). Because of the way the survey question was asked ("What's the most you would pay for one ticket [2 throws]?"), student solvers had to construct a cumulative frequency distribution rather than deal with only the raw frequency for each specific price. The estimated income based on the results of the sample then had to be extrapolated to the school population.

Best ticket price. Figure 5 shows the percentages of college and sixth-grade students who mentioned, attempted, and accurately estimated the best ticket price. All of the college students and all but one sixth grader mentioned ticket price, but only three (21.4%) of the sixth graders and eight (50%) of the college students attempted any calculations to determine the best ticket price. The others appeared to choose $1 as the best ticket price on the basis of the highest raw frequency (21 people, as shown in the survey chart). These individuals gave no indication that they considered the cumulative distribution or that they determined the sample revenue for any of the other ticket prices. Of those who did try to calculate the best ticket price, college students were largely successsful in that seven of eight arrived at the correct answer. Among sixth graders, only one student correctly solved this part of the problem.

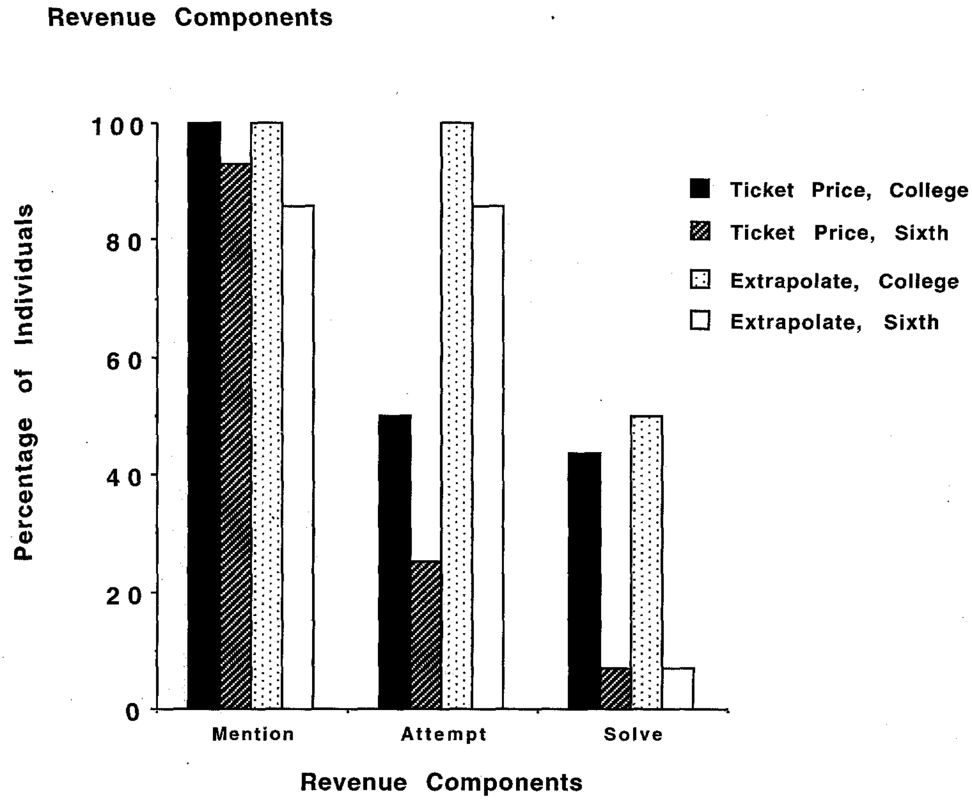


FIGURE 5 Percentages of participants including revenue components in Experiment 1.

Extrapolation to the population. The data in Figure 5 for extrapolation show that only half of the college students and only one of the sixth graders correctly solved this part of the problem. The low solution rate contrasts with the high rates of mentioning and attempting to determine income for the school population. All college students and 86% of the sixth graders mentioned and tried to figure out how much the whole school would pay. A frequent error for all students was to use the wrong number for the school population (i.e., 380, the total number of students enrolled in the school, instead of 360, the average number of students who were present on a given day). Also, as we noted earlier in our discussion of the error type data, formulation errors were common for sixth graders and college students when they were attempting to determine revenue to be raised from ticket sales. The nature of these errors, however, was different for the two groups. College students' errors were related to confusions regarding the appropriate ratio in going from the sample to the population, a problem indicative of inadequate understanding of the part–whole relation between the sample and the population. In contrast, the formulation errors made by the sixth graders suggest that many did not understand that a sample could be used to estimate population parameters. These students did not use the part–whole relation of the sample to the population. Instead, they indicated that everyone in the school would come, said that half the school would come (an estimate that was mentioned in the video by Chris before he did his survey), or added the frequency data from the sample and used that total as the estimated revenue from ticket sales.

Expenses

The plan elements for expenses reflect fixed and variable costs. Fixed costs included the cost of renting the dunking machine and the pool and the costs for setting up the pool. Variations in costs emerge because these depend on the selected method of pool filling and emptying. The method of filling also had implications for the number of days the pool had to be rented.

Fixed costs. Eighty-six percent of the college students and 65% of the sixth graders mentioned the cost for the dunking machine. In addition, all students mentioned and attempted to determine what the fixed costs for the pool would be (i.e., setup charges). Less than half the sixth graders, however, and 69% of the college students accurately determined the fixed costs for the pool. Errors were due to omission of one or more of the relevant costs and to incorrect determination of the hourly charges for setting up the pool.

Variable costs. There was variation in the pool-filling and pool-emptying plans that individuals chose and in the costs associated with them. For example, pool-filling and pool-emptying plans that involve using the pool store have costs associated with them, whereas using the school hose or the fire department to fill the pool results in no costs. The school hose takes longer than the other methods, however, and the fire department method is risky because they might be unavailable due to a fire alarm.

First, we report the results for the pool-filling plans and then for the pool-emptying plans. Among the college students, 50% chose to use the fire department to fill the pool; 31% chose the pool store plan (an expensive plan that has no risk); 13% chose a combination plan, that is, buy one load of water from the pool store and fill the rest of the pool with the school hose; and 6% chose some other type of plan. Among sixth graders, 29% selected the fire department plan, 50% chose the pool store plan, 7% chose the pool store and school hose combination plan, and 14% chose some other plan.

In pool-filling plans for which there were associated expenses, students were very likely to mention and attempt to determine the costs. All of the nine college students proposing plans that had expenses associated with them mentioned and attempted to determine these costs. Only 33%, however, accurately solved the expenses for their pool-filling plan. Similarly, for the sixth graders—nine of whom selected pool-filling plans that had expenses—eight mentioned and attempted to solve for these expenses. Only one sixth grader, however, accurately solved the expenses. Errors in solving for costs were most frequently associated with plan element omissions.

Pool-emptying plans were specified by fewer students than were pool-filling plans. Among college students, 11 of the 16 decided to use the public works truck to pick up the water, incurring no expenses; the other 5 students did not indicate how they would empty the pool. Among the sixth graders, only 3 students decided to use the public works truck, 7 mentioned that they needed to think about how to empty the pool and what it would cost but did not go any further with it, and the remaining 5 students did not mention how they would empty the pool at all.

Feasibility: Time and Risk Factors

In *The Big Splash*, time is one aspect of feasibility with which students have to deal, regardless of the specific plan. Risk is a second aspect of feasibility, but it is relevant only if certain plans are selected.

With respect to time, one needs to know if there is enough time to implement a particular plan successfully. Fewer than half of the college or sixth-grade students mentioned or attempted to determine if they had enough time to execute their pool-filling plan. For the pool-emptying plan, virtually no one considered time, possibly because those who considered a pool-emptying plan selected the public works truck. The video did not include information about how long it would take to empty the pool using this method.

The time aspect of plan feasibility may have been overlooked by a substantial number of individuals because the challenge did not explicitly mention it. In the video, however, time was mentioned in connection with several plan components for setting up the dunk tank and should have been considered in calculating a number of the expenses. Time may function as an embedded subgoal, and the low incidence of considering it may be comparable to previous results in which we have found that embedded goals—especially those arising in response to an obstacle—are not dealt with frequently (CTGV, 1994; Goldman & CTGV, 1991).

Another aspect of feasibility is risk: Is the plan too risky to count on? Information in the video indicated that the fire department could not fill the pool if they were called out on an alarm. Of the seven college students and four sixth graders who chose the fire department to fill the pool, only three of the college students and three

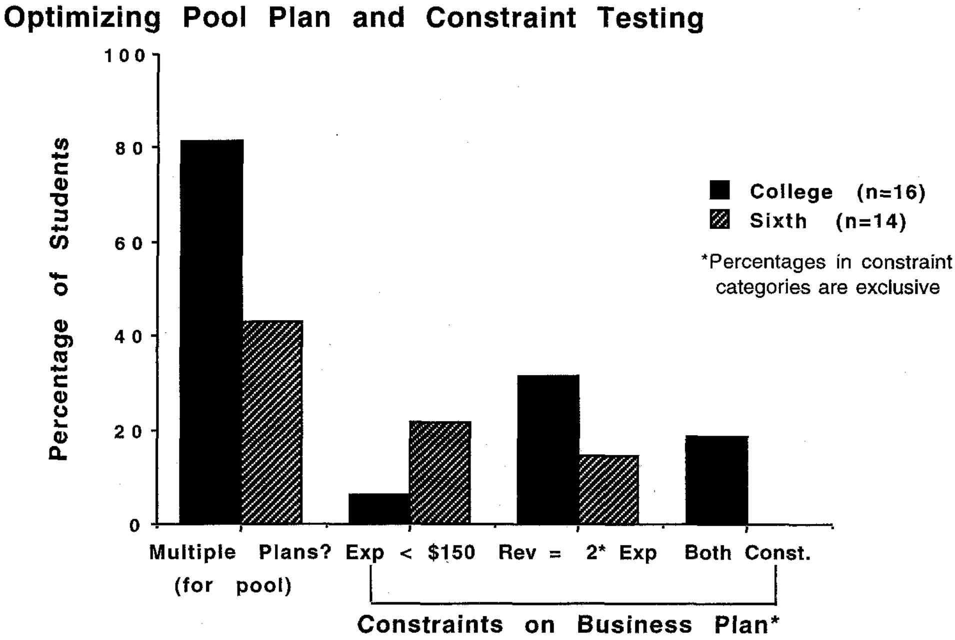


FIGURE 6 Percentage of participants including optimization and constraint testing elements in Experiment 1.

of the sixth graders explicitly considered the likelihood that the fire department would be unavailable on the day of the fair due to a fire alarm.

Optimization and Constraint Testing

The data reported thus far on plan elements have been for the plan that the students selected as the best business plan, if they gave more than one. In previous work, Goldman and CTGV (1991; see also CTGV, 1994) reported the tendency for students not to consider more than one plan even though the one they had selected might not have been optimal. In fact, in solving a trip-planning adventure, the percentages of problem solvers who continued problem solving in an attempt to optimize the plan were relatively rare. Figure 6 suggests a similar trend in the business plan case but only for the younger students: 81.3% of the college students did consider multiple plans but only about 43% of the sixth graders did.

The data in Figure 6 also indicate that a large number of students might not have been able to convince Ms. Stieger, Chris's school principal, that their business plans met the constraints she had set forth (expenses under $150 and revenue twice as much as expenses). Only 56% of the college students and 36% of the sixth graders tested their plans against one or both constraints.

Summary of Experiment 1

Finding a solution to *The Big Splash* requires the use of statistical inference; students were not very good at this, even those who were enrolled in college. When they were inaccurate, they were misled by high raw frequencies, they failed to construct the cumulative distribution called for by the problem, and they had difficulty formulating the appropriate ratio needed for extrapolation. In addition, some sixth graders lacked the fundamental notion that sample data could be used to estimate population parameters.

Plan feasibility was not carefully tested with respect to time constraints; students focused on expenses. The need to consider time is actually embedded in costs for the pool. The low frequency of attending to this subgoal may be due to its embedded status, consistent with our previous findings on embedded goals (CTGV, 1994; Goldman & CTGV, 1991).

Although the business plan domain provides explicit constraints against which to test solutions, the majority of students failed to do so. Despite the focus on costs at the pool plan level, most students did not test whether their business plans, in fact, met Ms. Stieger's overall constraints.

In the business plan domain, in comparison with trip planning, students were somewhat better able to coordinate different pieces of information with the appropriate plan element—perhaps because of the salient differences among the plan options.

Major sources of differences between college students and sixth graders were in the number of attempts and correct solutions, in the tendency to consider multiple plans, and in the breadth of exploration of the solution space. There were not any particular differences in their skill at generating the subproblems that needed to be solved.

EXPERIMENT 2: FIFTH-GRADE DYADS' BUSINESS PLAN SOLUTIONS

Experiment 1 indicated that complex problem solving in the business-planning domain is as difficult for students as complex trip planning. The plan element analysis indicated that there were major gaps in understanding the statistical concepts associated with revenue. In contrast, the sixth graders seemed to understand the concept of expenses, but the errors indicated the need for additional monitoring of the solution process. Particularly interesting was the failure to consider alternative business plans. This replicates the finding from trip planning that students tended to stop short of comparing alternative plans, even when they had not chosen the fastest plan as their own solution (CTGV, 1994; Goldman & CTGV, 1991).

Individual problem solvers in Experiment 1 tended not to articulate their justifications for particular actions or their interpretations of solution outcomes. Our goal in Experiment 2, therefore, was to understand more fully the nature of the reasoning processes associated with different solution outcomes. Consequently, we had dyads solve *The Big Splash* cooperatively. We anticipated that the problem-solving protocols produced by dyads would contain greater information about reasoning and decision-making processes than was available in protocols produced by individuals. We also examined the contributions made by each member of the pair to identify the patterns of interactions associated with more successful problem solving.

Method

Materials

The materials for Experiment 2 were the same *Jasper* adventure (*The Big Splash*) and accompanying materials as were used in Experiment 1.

Participants

Participants were 34 fifth graders (15 boys and 19 girls) drawn from a public school system in a suburban township located in the Southeast. Percentile scores on the math component of the Tennessee Comprehensive Achievement Program ranged from the 74th to the 99th percentile ($M = 88$). Eleven of the students scored at the 90th percentile or better. Dyads were formed to match percentile scores and gender, but one dyad was mixed gender.

Design and Procedure

Each dyad participated in two 45-min sessions in a room outside their classroom. Participants were asked to respond orally, and all responses were videotaped. In the first session, dyads were initially presented with *The Big Splash*. During the video presentation, participants were given paper and pencils in the event that they wanted to take notes, although they were not instructed to do so. After watching the video, each dyad received the following general instructions:

> For the rest of the period, we will be focusing on the challenge. In a few moments, I will be giving you an information sheet with all the important information from the video. First, though, I want to talk to you about working together. I want you to concentrate on working as a team. Think out loud as much as possible so we can tell what you're thinking and so your partner understands what you are doing. To be a good team, people need to talk to each other. So, listen closely to your partner and what he or she says. If your partner says something you don't understand, ask him or her to explain, or if your partner says something that you don't agree with, tell your partner and explain why you do not agree.

Following these general instructions, each dyad was presented with two tasks. The first was a 10-min brainstorming task in which students were asked to think about and construct a list of the important things one needed to do in order to solve the challenge. Instructions for this task were as follows:

> The first task I want you to do is think about the challenge without making any calculations. Specifically, I want you both to spend a few minutes thinking about and writing a list of the most important things Chris had to think about in preparing his business plan.

After the brainstorming session, the experimenter indicated that two important factors that Chris had to consider in developing his business plan were his estimated expenses and his estimated income. The experimenter then proceeded to solve the income side of the problem, prefacing this action with the following: "What I'd like to do now is show you how Chris figured out what his income would be, and then later, your task will be to figure out what his expenses and profit would be." After the experimenter explained the income portion of the problem, participants received the aforementioned fact sheet and were instructed to begin the next phase of the study—the problem-solving task. In this task, the dyad spent the remaining part of the session, or about 15 min, solving the *Jasper* challenge. In the second and final 45-min session, participants were instructed to continue to work on the *Jasper* challenge.

Results and Discussion

The dyads' problem solving was first analyzed using two of the same analyses—the error and solution-space analyses—that were conducted on the think-aloud protocols generated by the individuals in Experiment 1. Two coders independently scored 25% of the protocols. Coder agreement exceeded 90%. Disagreements were

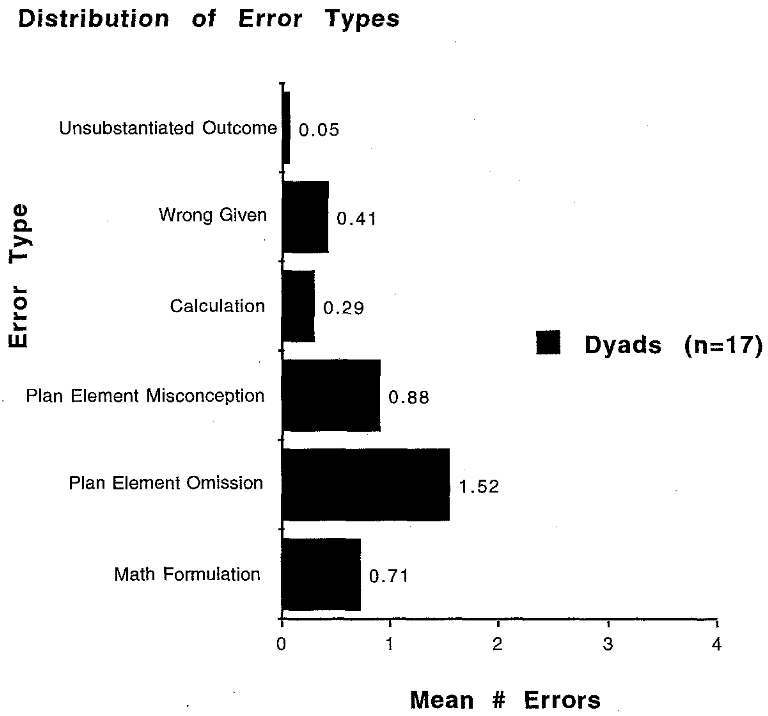


FIGURE 7 Distribution of error types in Experiment 2.

resolved through discussion. A third set of analyses examined the reasoning that occurred in the dyads with respect to goals, arguments, their interrelations, and their relations to calculations and solution-space coverage.

Types of Errors

The mean number of errors made by dyads was 3.94. The distribution over the types of errors is shown in Figure 7. There were significant differences among the types, $F(6, 16) = 8.31$, $MSE = 0.92$, $p < .01$. The Newman–Kuels procedure indicated that all pairwise comparisons were significant. As in Experiment 1, the most frequent type of error was plan element omission, accounting for 39% of the errors. The next most frequent types of errors were plan element misconceptions (22%) and math formulation errors (18%). Also consistent with Experiment 1, relatively few dyads produced incorrect solutions as a consequence of making computational errors (7% of the errors).

Although there were procedural differences between Experiment 1 and 2, it is still interesting to compare these results to those obtained from the older participants in Experiment 1. Dyads made fewer errors than sixth graders working alone but more than college students. Plan element omissions were lower for the dyads than for the sixth graders but about the same as the percentage for the college students. Finally, dyads produced a higher percentage of errors due to misconceptions but about the same percentage of formulation errors. We speculate that dyads may have been less likely than individual children to omit plan elements because dyad members monitor their partners' plans and, when they can, contribute information about missing plan elements.

Plan Elements in the Solution Space

To describe both the extent to which dyads explored the solution space and their success in doing so, we examined the plan elements related to expenses that they included in their solutions. As in Experiment 1, plan elements were scored with respect to three levels of inclusion. That is, we determined if a plan element had been mentioned, if an attempt had been made to solve the element, and if the element had been successfully solved. For purposes of these analyses, we examined the plan for estimating expenses that dyads indicated was their best.

Expenses

Fixed costs. As discussed earlier, the fixed costs associated with the problem are the cost of the dunking machine (a flat fee that is given in the problem) and the costs associated with procuring the pool (i.e., the daily rental charge for the pool and the labor costs for assembling and disassembling it).

The fixed cost for the dunking machine was mentioned by all but one pair of students. In addition, all pairs mentioned and attempted to compute the fixed costs for the pool, and the majority of these students were successful. Relative to the data from Experiment 1, a greater number of fifth-grade dyads than individuals (i.e., sixth graders) mentioned the cost of the dunking machine (94% and 65%, respectively) and successfully solved the fixed costs for the pool (59% and 43%, respectively).

Variable costs. Recall that a major aspect of the problem space involves generating and evaluating plans for filling and emptying the pool. The available options differ in terms of cost and feasibility with respect to time and risk.

All dyads mentioned that they had to consider expenses for filling the pool. Fifty-nine percent of the pairs, however, selected a pool-filling plan that was free (i.e., water from the fire department or water from the school hose) and consequently did not have to compute any costs for the water. The remaining 41% of the dyads chose to buy some amount of water from the pool store and attempted to compute how much that would cost. Only one pair, however, produced a correct answer. The most common errors were plan element omissions, for example, omitting the cost of a second load of water or the charge for mileage. Formulation errors were also common. In particular, dyads had difficulty determining how to use odometer readings to calculate mileage.

In terms of emptying the pool, only one option (pool store truck) involves any expense. Seventy-six percent of the dyads mentioned a plan for water disposal, but none of these students selected the pool store truck as their means for removing the water. Hence, there was no need to consider cost.

With respect to pool-filling expenses, the results for the dyads and individuals in Experiment 1 were very similar. Virtually all students mentioned the water-filling costs and, when appropriate, attempted to determine total costs. With rare exception, however, the total costs were incorrect, in large part because students failed to include all relevant expenses. In contrast, the dyad and individual data for pool-emptying expenses are quite different. Dyads selected methods that were free; the majority of individuals selected a method with associated costs but did not consider what those costs would be.

Feasibility: Time and Risk Factors

As noted in Experiment 1, plan feasibility with respect to time constraints was not really considered by individuals working alone. We speculated that the low frequency of attending to time may have been due to its embedded status in the planning net; time is not explicitly mentioned in the challenge and only emerges as a subgoal in calculating expenses.

The majority of dyads mentioned and attempted to solve the time subproblem. About twice as many dyads as individuals working alone mentioned filling time (88% vs. 43%, respectively) and attempted time calculations (59% vs. 29%, respectively). The success rate of these attempts to solve time, however, was low for dyads, as it had been for individuals in Experiment 1. Dyads who attempted to determine how long it would take to fill the pool if they hired the pool store were most often unsuccessful because they omitted the time associated with filling and transporting the second load of water. Dyads who attempted to determine how long it would take to fill the pool with the school hose were most often unsuccessful because they did not correctly formulate the problem, which involves using rate information and proportional reasoning.

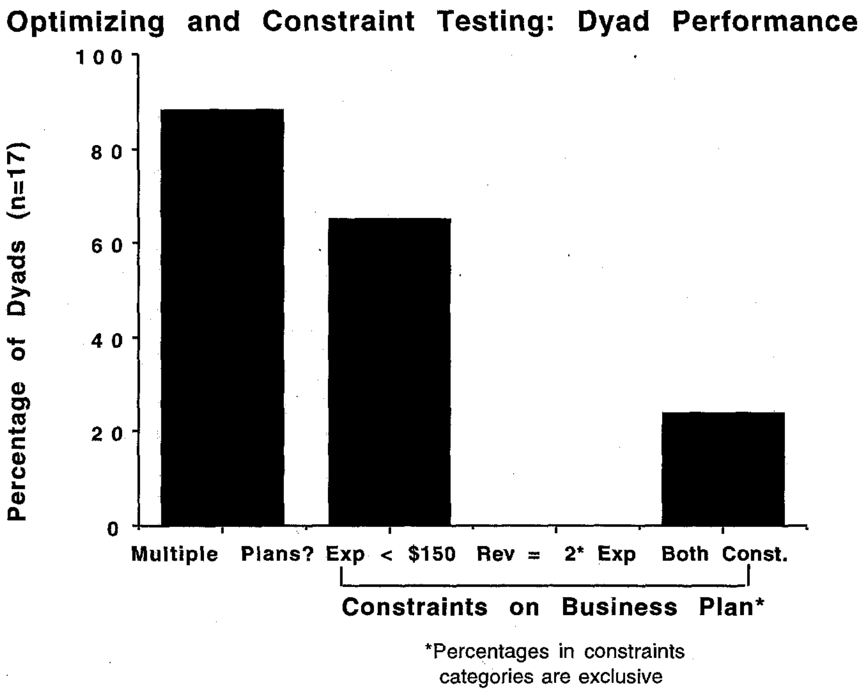


FIGURE 8 Percentage of participants including optimization and constraint testing elements in Experiment 2.

The data on pool-emptying time indicated that only two dyads attempted to find out how long it would take to empty water from the pool. The low salience of this element for the dyads replicates the finding for individuals.

Optimization and Constraint Testing

Figure 8 shows that the overwhelming majority of dyads (15, or 88%) attempted to optimize their plans by considering more than one way to fill the pool. With respect to the performance of individuals, these dyads performed at the level of the college students and considered multiple ways to fill the pool about twice as often as the high-achieving sixth graders. Because of several procedural changes between the two experiments, we cannot attribute this difference solely to working in dyads. One advantage of dyads over individuals, however, may be that dyads are better able to monitor each other's problem solving and can remind one another of important problem constraints.

Figure 8 also shows the percentage of dyads who considered the financial constraints. The overwhelming majority of the dyads (88%) considered one or both financial constraints. As with the time constraints, this level of performance compares very favorably with the performance of college students in Experiment 1 (Figure 6).

The results of the plan element analyses suggest that dyads are very likely to consider multiple-solution plans and to test the feasibility of their chosen plans, both with respect to the time needed to execute these plans and with respect to the constraints on expenses and profit. In addition, the analyses suggest that, relative to the performance of sixth graders in Experiment 1, dyads mentioned and attempted to solve more of the expense elements of the solution space. Again, these outcomes are consistent with the idea that important functions for the second member of the dyad are to monitor the problem solving and to share knowledge that advances exploration of particular solution plans—as well as knowledge of alternate plans and problem constraints. This is not to say that individuals cannot attend to their own problem solving in this way, but it may be that, in problem-solving situations as complex as *The Big Splash*, they do not do so as frequently or as effectively. It is certainly possible that alterations to the task context that reduce the complexity of the situation might lead to more individuals monitoring their problem solving; this would be an interesting experiment to conduct.

The enhanced performance of dyads relative to individuals was anything but a foregone conclusion. The interaction among the members of the dyads might well have impeded problem solving (cf. Goldman et al., 1992; Salomon & Globerson, 1989). The key issue concerns how the dyads interact in the problem-solving process (e.g., Webb & Palincsar, 1996; D. Wood, Wood, Ainsworth, & O'Malley, 1995). In the next section, we examine the problem-solving process more closely. We first consider the relations among three major processes involved in complex problem solving: goal generation, reasoning about those goals, and the application of mathematical calculations to meet those goals. We then examine patterns of interactions associated with effective problem solving.

The Goals, Reasoning, and Application Framework

We expanded each element of the planning net representation by "unpacking" it into the more detailed series of goals and subgoals constituting each element of the solution space depicted in Figure 2. In so doing, we generated 125 goals and subgoals, hereafter referred to as *goals*. A portion of the expanded planning net structure is presented in the Appendix. This expansion allowed us to capture problem-solving processes at the level of detail reflected in the dyads' conversations. We first report some basic descriptive information about goals, arguments, and mathematical calculations present in the dyadic protocols of the problem-solving portion of the experiment. We then report on the contingency relations among them.

Goals

Examination of the data revealed that dyads generated 237 goals, 168 of which were associated with the goal structure—140 correct and 28 incorrect. (Incorrect goals were within the goal structure but not stated accurately.) In addition, 69 goals were stated that were outside the goal structure, such as "to know what teachers needed to be dunked." The data thus indicate that 71% of the goals generated by the dyads were associated with the goal structure of the solution space proper—83% of which were correct. There was considerable variation across dyads with respect to the total number and proportion of goals generated per dyad. The number of goals stated by a dyad ranged from 4 to 29, with the proportion of correct goals ranging from .29 to .89.

The goals in the goal hierarchy were categorized as high, medium, and low. High-level goals such as "estimate profit" and "select pool plan" were global in nature and referred to the three highest levels of the hierarchy. Medium-level goals such as "determine cost for filling the pool using the store option" were drawn from the middle three levels of the hierarchy. Finally, goals of the lowest three levels, such as "determine cost of filling pool" and "determine mileage from pool store," represented the most specific goals. In general, the lower levels of the hierarchy reflected greater specificity. The data indicated that students mentioned approximately the same percentage of high and medium goals (46% and 44%, respectively), but rarely generated low-level goals (10% of total goals). Furthermore, this pattern of results obtained from the problem-solving sessions was consistent with similar (but not reported) analyses of goals mentioned during the brainstorming session.

Arguments

An argument was defined as a conclusion supported by one or more reasons. Argument soundness was evaluated on the basis of two criteria: the relevance of the reason to the conclusion and reason acceptability (i.e., whether the reason supported the conclusion; Angell, 1964). A total of 439 arguments were generated, with 339 (or 77%) being judged as sound. As with goal statements, the data indicate considerable variation in performance across dyads with respect to the number ($M = 20.0$, range = 1–46) and proportion of sound arguments generated per dyad ($M = .76$, range = .48–1.00).

The content of the dyads' reasoning was classified into one of five types of arguments:

1. Goal-related arguments involved reasoning about goals, for example, "He needed an alternate plan in case they were on a fire so they maybe have to get the water from ... the pool people."
2. In fact-related arguments, dyads reasoned about facts or inferences pertaining to the facts of the story, for example, "the dunking machine is $25 a day. ... No, they said they would give it to him for free."
3. Calculation-related arguments were those in which calculations were stated in the form of an argument, for example, "It's $7 an hour and 2 hours to set up, so that's $14."
4. Examples of metacognitive-based arguments were "It doesn't say. Yes [it does]. We've got that already." Students also reflected about the problem or their progress through the problem.
5. The remaining category was irrelevant arguments. These were comprised of opinions or statements beyond the scope of the solution space, for example, "So that's a lot of money," which was said in the context of the estimated revenue of $270.

Analysis revealed that the distribution of sound arguments across type was as follows: (a) goal related—38 (11%); (b) fact related—116 (34%); (c) calculation related—117 (35%); (d) metacognitive—34 (10%); and (e) irrelevant—34 (10%). The data thus indicate that 69% of the arguments involved supporting or disputing facts and calculations. Arguments were also examined with respect to whether they supported a given claim or whether they opposed a given claim (i.e., were counter-arguments). Of the 339 arguments generated, 230 (68%) were supportive, whereas 109 (32%) were counterarguments.

Mathematical Calculations

As indicated in the error analyses, calculations were not often sources of solution errors. The mean proportion of calculations producing correct solutions was .66, but the range was large—.33 to 1.00. Likewise, there was substantial variability in the number of computations attempted: The range was 6 to 17, with a mean of 9.29.

Interrelations Among Goals, Arguments, and Calculations

In this section, we consider data pertaining to the interrelations of the previously described three components. The primary question addressed here concerns the coherence and directedness of dyads' thinking as it relates to particular goals in the goal structure. In other words, do the co-occurrences and proximities among goals and arguments or goals and calculations suggest a purpose or organized direction for the problem solving? To address this question, we examined the co-occurrence relations between goals and arguments and between goals and calculations (e.g., Is a goal followed by an argument justifying it? Does a calculation relevant to a stated goal follow that goal?). We determined the conditional relations among four pairwise linkages: (a) Given a goal, what was the occurrence of an argument related to the goal? (b) given an argument, what was the occurrence of a goal related to the argument? (c) given a goal, what was the occurrence of a calculation relevant to meeting the goal? and (d) given a calculation, what was the occurrence of a goal that followed from the outcome of the calculation? For each dyad, the number and proportion of these contingencies were calculated. The conditional probability data for each dyad are provided in Table 1.

Goal-to-argument linkages. An argument may serve a number of functions in relation to a goal—for example, as justification or evaluation of the goal. Theoretically, the more arguments stated in relation to goals, the more likely that dyads would be moving successfully through the problem space. There are two ways to view goal–argument co-occurrences. The first examines arguments condi-

TABLE 1
Summary of Goal Statement and Argument Linkages and Goal Statement and Calculation Linkages

Dyad	*P(Arg/Goal)*	*P(Goal/Arg)*	*P(Calc/Goal)*	*P(Goal/Calc)*
1	.86	.86	.91	.99
2	.20	1.00	.99	.99
3	.33	.13	.50	.00
4	.69	.61	.50	.88
5	.75	.21	.67	.33
6	.71	.61	.71	.86
7	.29	.83	.50	.99
8	.20	.83	.60	.99
9	.54	.83	.82	.55
10	.29	.57	.33	.50
11	.33	.80	.43	.99
12	.44	.43	.67	.75
13	.50	.55	.29	.99
14	.67	.15	.00	.00
15	.78	.50	.43	.17
16	.99	.83	.99	.80
17	.99	.69	.71	.88
M	.56	.61	.59	.69

Note. Arg = argument; Calc = calculation.

tionalized on the occurrence of a goal. That is, given the occurrence of a goal statement, did the particular dyad state one or more arguments related to that goal? For example, Dyad 1 generated 18 goals, 12 of which have at least one argument stated in relation to that particular goal. The following sequence of statements from Dyad 1's protocol illustrates a goal to argument linkage: "We need to see what everything costs. It's only for a day so the dunking machine will be for $25.00."

The data in Table 1 indicate that, for the majority of dyads (10 out of 17), the probability of a goal having related arguments was greater than .50. The overall average was .56. These data indicate that the majority of goal statements were followed by a related argument.

The second way to look at goal–argument links is to examine goals given the occurrence of argument statements. In terms of problem solving, this conditional relation differs from the first in the following way: Imagine a problem-solving situation in which many arguments and counterarguments are made but there are no obvious purposes for engaging in the debate. By examining the occurrence of a related goal, given the occurrence of an argument, we assess the degree to which argument statements were purposeful in directing the problem solving. The conditional relation of goals given arguments differs from the first conditional relation (i.e., arguments given goals) in that it looks at whether arguments led to the formation or realization of new goals. For example, Dyad 2 generated 42 sound arguments, 36 (86%) of which were related to goals. Thus, for this dyad, most reasoning occurred around goals. The second column of Table 1 presents the proportion of goal statements, given the statement of an argument. The mean proportion of arguments leading to new goals is .61, and for 13 dyads more than half of the argument statements were purposeful. In Dyad 1, for example, one student stated:

> If they [fire department] were out on call, then they would have to get it [water] from somebody else, so let's try to average out [calculate] what a second person would be [cost] if they couldn't get water from the fire people.

Also, Dyad 9 jointly stated "That's $158. ... Oh!" followed by this comment by one member of the dyad: "Let's find something cheaper" ($158 was greater than the $150 limit).

The data on goal–argument linkages indicate that dyad reasoning is coherent with respect to goals and arguments. Goals tend to be followed by related arguments, and problem solving does not tend to be characterized by purposeless argumentation. Rather, argumentation leads to new goals.

Goal-to-calculation linkages. A similar logic applies to goal–calculation linkages. The higher the probability that a calculation occurred, given the statement of a goal to which the calculation was relevant, the more goal directed the problem solving and the greater the likelihood of success. For example, Dyad 1 generated 11 goals, 10 of which were followed by at least one calculation. The results presented in the third column of Table 1 indicate that the dyads were reasonably goal directed in that the mean proportion of cases in which a calculation occurred after the statement of a related goal was .59. For 12 dyads, the probability was .50 or greater that a goal statement was followed by at least one calculation. An example of a goal to calculation linkage is the following: "And how much to set it up? [students perform calculation]. And the cost is $7.00 an hour, so this would come to $14.00."

The second contingency examined is the probability that a calculation performed to meet a specific goal led to the statement of a new goal. Computationally, this is the occurrence of a goal given the occurrence of a calculation. The data provided in Table 1 reveal that the mean proportion of goals stated, given a calculation, was .69, and that, for 13 of the dyads, .50 or more of the outcomes of calculations set up the conditions for stating new goals. Thus, for the most part, calculations occurred in the service of active goals, and they advanced problem solving in that they tended to lead to new goals being stated.

Correlational Analyses of Goals, Arguments, and Calculations and Solution-Space Elements

The analyses just discussed indicate that, for the majority of dyads, there is coherence or direction to the thinking and problem-solving process. We were also interested in the degree to which the various indexes of process and problem solving were interrelated. For example, was there a relation between the number of sound arguments and successful performance on calculations? Or was there a relation between the number of goals stated and successful performance on calculations? To investigate these questions, we conducted correlational analyses among the process measures derived from the protocols (e.g., number of goals, number of correct computations, the conditional measures) and a global measure of the extent of the solution space that each dyad explored. The latter was derived from the analysis of the plan elements in the solution space and had a possible maximum of 10.[1] The correlational analyses speak to the issue of "individual" differences among the dyads in that high correlations among the process measures indicate the degree

[1]Each dyad received one point for each of the following elements of the solution space that they attempted to solve: dunking machine, labor for pool, pool rental, costs to fill the pool, time to fill the pool, cost to empty the pool, time to empty the pool, if expenses were greater than $150, if estimated revenue was less than twice the estimated expenses, and if multiple plans were considered. The distribution of actual scores was 5 to 10 over the 17 dyads, ($M = 7.4$).

TABLE 2
Correlational Analyses for Indexes of Problem Solving in Experiment 2

	SA	*CC*	*SA/CG*	*CG/SA*	*CC/CG*	*CG/CC*	*EL*
CG	.54*	.66**	.20	.41	.24	.38	.19
SA		.69**	.85**	−.02	.29	.10	.66**
CC			.59*	.37	.63**	.24	.45
SA/CG				−.11	.22	−.11	.56*
CG/SA					.59*	.81**	−.28
CC/CG						.44	.12
CG/CC							.45

Note. CG = correct goals; SA = sound arguments; CC = correct calculations; SA/CG = sound arguments, given correct goals; CG/SA = correct goals, given sound argument; CC/CG = correct calculations, given correct goals; CG/CC = correct goals, given correct calculations; EL = no. of elements in the solution space (max. = 10).
$^*p < .05$. $^{**}p < .01$.

to which indexes of sound argumentation and successful problem solving tend to co-occur in the same dyads. Table 2 contains the correlation matrix.

We first consider the correlations among the independent occurrences of the various indexes of the problem-solving process: frequency of stating correct goals, frequency of stating sound arguments, and number of calculations correctly performed. There were significant correlations among all three of these measures, (r = .54 for correct goals with sound arguments, r = .66 for correct goals with correct calculations, and r = .69 for correct calculations with sound arguments, ps < .05). These findings indicate that dyads who generated more goals also tended to generate more sound arguments, and dyads who generated more goals also tended to apply their mathematical skills successfully.

The correlations among these independent measures and the conditional measures further specify the nature of the problem solving. The relevant data are the following: (a) the greater the number of sound arguments, the more likely they were conditional on a relevant goal statement (r = .85); (b) the higher the frequency of correct calculations, the greater the tendency for them to be conditional on a related goal (r = .63); (c) the frequency of correct calculations was significantly correlated with the frequency of sound arguments conditional on a goal (r = .59); and (d) frequency of correct calculations conditional on a goal correlated significantly with frequency of sound arguments conditional on a goal (r = .59).

The absence of significant correlations between the conditional measures and the frequency of stating goals indicates that it is not "throwing out goals" that drives coherence. Rather, coherent problem solving is characterized by sound reasoning about goals in conjunction with correct application of mathematics skills in meeting those goals. Coherent problem solving was also predictive of successful performance as indexed by the extent of the solution space that was attempted. Two process indexes, sound arguments and sound arguments conditional on goals, correlated significantly with the solution-space measure ($r = .66$ and $r = .56$, respectively). Correct calculations and the likelihood that correct calculations led to new goals were correlated at nonsignificant levels ($r = .45$ in both cases) with attempting more of the elements in the solution space.

The correlational data indicate that dyads were more successful in their problem solving only if goals called forth sound reasoning and accurate calculations. Merely doing correct calculations was insufficient to produce coherent solutions to the problem, even if those calculations led to stating new goals. Hence, goal-directed reasoning and application of mathematics skills are important. This conclusion is consistent with previous data collected in the context of *Jasper* problem solving: Students' problem solving was hampered by difficulties identifying what data were relevant to which goals rather than in generating the goals or doing isolated computations on the data (CTGV, 1993, 1994).

Reasoning Within the Dyads

The solution-space data showed that, overall, dyads attempted more of the solution space than did individuals working alone. The correlational analyses of the dyads' goals, reasoning, and application of mathematics skills suggested that the dyads differed with respect to the coherence and directedness of their problem solving. Furthermore, those dyads with higher levels of sound reasoning in the context of goal-directed solutions attempted more of the solution space. In this section, we consider data relevant to two possible explanations for these effects.

First, working together may have engendered a situation in which each member of the pair tried to be explicit about his or her thinking, verbalizing for the benefit of one's partner exactly what was being done and why. The act of explaining to the other may have enhanced problem solving in and of itself. The function of these argument statements was to explain the speaker's thinking to the other member of the pair. A second explanation for the foregoing pattern of results resides in a second function of working in dyads. We suggested in the introduction that one possible benefit of working in dyads is that the members can monitor each other's problem solving. Arguments that show the monitoring function would reflect a reaction by one member of the pair to a proposal made by the other member.

There is some evidence that argument statements served both explanation and monitoring functions in the dyads in this study. The data indicate that 202 of 339, or 60%, of the sound arguments were individually generated, and 137 of 339, or 40%, were generated by the second member of the pair. In the first case, arguments are being supplied by the selfsame individual proposing a particular goal or calculation. That is, they explain the speaker's actions. In the second case, the argument generation serves the monitoring function and involves both members of the dyad. We refer to these as *dyad-generated arguments*.

The dyad-generated arguments were divided into pro arguments and con arguments. Dyad pro arguments consisted of the statement of a claim supported by a reason—with each member of a dyad stating one part. Thus, the reason, stated by Member 2, "They'd have to drive back and fill it up again," provides support for the claim of Member 1: "So they'd have to make two trips." In dyad con arguments, one partner disagreed with the other, typically by generating a counterclaim and reason in opposition to the stated claim. For example, Member 1's statement, "The fireman said the dunking machine rental was free," was opposed by Member 2's counterargument, "No, [dunking machine rental is not free] because it says here [fact sheet] dunking machine rental is $25 a day."

Examination of the data revealed that 45 of 137, or 33%, of the dyad-generated arguments involved one partner supporting the other, whereas 67% of the dyad arguments consisted of those in which the participant disagreed with his or her partner. The majority of dyad-generated arguments functioned to monitor and redirect the course of problem solving, consistent with the findings of Howe, Tolmie, Greer, and Mackenzie (1995). Classification of these arguments, using the classification scheme discussed earlier, revealed that 53% of the dyad-generated arguments involved disputing the "facts" that the other had proposed. Only 21% were about calculations, and 13% were about metacognitive issues. These results indicate that an important dyad monitoring function involved extracting accurate data from the story and establishing its relevance to the goal.

The analyses of the dynamics of the argumentation in the dyads revealed that reasoning processes involved both explanation of one's position and disputation of the other member's problem solving. The overall effect of these interactions was to encourage broader exploration of the solution space for expenses and to assist with determining which data were accurate and relevant to which goals. This enabled dyads to more successfully attempt and solve a number of the elements of the solution space.

Summary of Experiment 2

The dyads' problem solving tended to be relatively complete with respect to exploring elements of the solution space for expenses, especially with regard to optimization and constraint consideration. Approximately 75% of the dyads considered multiple plans and one or both of the financial constraints on the business plan. The goals, reasoning, and application analysis of the dyads' problem solving indicated that the more coherent the problem-solving process, the more solution-space elements considered. Characteristics of more coherent problem solving were sound reasoning about goals in conjunction with the correct execution of appropriate computational procedures on the relevant data. Sound reasoning involved individual members of the dyad offering explanations of claims they made as well as counterclaims offered by the other member of the dyad. This dialogic process enriched and broadened the extent of the problem-solving process.

GENERAL DISCUSSION

One of the primary concerns of this research is the development of a methodology for characterizing and describing problem solving in very complex situations. For the complex problem situations studied in this research, we adopted a technique that used modified planning net representations to specify the major elements of the space and the relations and constraints among them. We then mapped information obtained from think-aloud protocols (Experiment 1) or dyadic interactions (Experiment 2) onto the elements and relations in the planning net representation. Our investigation in this context focused on the business-planning domain. In developing business plans, it is necessary to consider the relation between revenue and expenses, attempting to optimize profit while still delivering a feasible plan.

The data from Experiment 1 indicate that, although college-age individuals are aware of many constraints on business plans, their problem solving does not always or accurately take these elements into account. In contrast to sixth graders, college students' solutions were characterized by higher frequencies of successfully attempting to deal with all plan elements. Among both sixth graders and college students, success was constrained by conceptual understanding of certain mathematical concepts (e.g., proportional reasoning), incomplete consideration of some of the financial constraints, and a tendency to overlook feasibility (i.e., timing) issues in their plans. Merely changing the domain from trip-planning to business-planning, in which constraints are more explicit, did not substantially alter our conclusion about the presence of little spontaneous consideration of multiple plans by middle school age students (e.g., CTGV, 1994). The business-planning domain did enhance multiple plan consideration by college-age students.

Complex problem solving of the type required to solve a *Jasper* adventure is a relatively new experience, even for high-achieving mathematics students. Their performance frequently reflected incomplete rather than inaccurate problem solving. As might be expected from this caliber of math student, few of their errors were due to calculation problems.

Experiment 1 provided information about the extent of the solution space considered by individuals working alone, but the think-aloud protocols did not contain sufficient information to understand fully individuals' reasoning processes. Experiment 2 was designed to make thinking more visible by creating a dyad situation with a need to communicate. Dyads tended to do a relatively thorough job of exploring different elements of the solution space for expenses: Almost all of them considered feasibility and optimization (multiple plans), although the success with which they dealt with these issues was far from perfect. For example, for those plans that involved costs, the problem solving of all but one dyad led to an inaccurate solution. The errors most often reflected omission of an element or formulating the mathematical relations inappropriately.

In addition to examining the extent of the solution space that was considered, we examined the coherence of problem solving via analyses of goals, arguments, calculations, and contingency relations among them. Our analyses of the reasoning processes indicated that dyad reasoning was coherent with respect to goals and arguments. Goals tended to be followed by related arguments, and the problem solving did not tend to be characterized by purposeless argumentation. Although the number of goals and arguments was highly variable, for the majority of dyads, there was coherence or direction to the thinking and problem-solving process. The contingency and correlational analyses indicate that coherent and successful problem solving is not predicted merely by the number of goals. What is predictive is sound reasoning about those goals and the correct execution of appropriately selected mathematical operations. In other words, what is important to successful problem solving is goal-directed reasoning and application of mathematics skills.

Dyadic interactions that supported goal-directed reasoning and application had two characteristics: explanatory reasoning and monitoring via pro and con counterarguments. Dyads differed with respect to the goal directedness of their solutions. The dialogic of the reasoning involved both explanations and countersuggestions. Both kinds of reasoning served a monitoring function and enhanced the extent of the solution space that was considered during problem solving.

Although there were a number of differences between the problem-solving situations in Experiments 1 and 2, it is noteworthy that the performance of the fifth-grade students in Experiment 2 looked similar to that of the college students in Experiment 1. This was true for the extent to which they explored the problem solution space and for the patterns of errors they made. The interaction data suggested that the explanation for the similarities across fifth-grade and college students may be in the degree to which members of a dyad can monitor the solution process and keep in mind the constraints and search space relevant to the problem. Members of the dyad may fluidly adopt different roles in problem solving as they switch between being listener and speaker in the verbal interaction. Listening from the perspective of checker, reflector, or coach (Kagan, 1992) may be an important function of the nonspeaking dyad member. These interpretative speculations are consistent with potential explanatory mechanisms for the value of peer interaction recently suggested by Webb and Palincsar (1996) in their review of group processes in the classroom.

What are the implications of these findings vis-à-vis mathematical problem solving in classroom settings? One of the goals of the *Jasper* series is to provide meaningful problems for students to solve that capture the intricacies of real-world mathematical problem solving. By this we mean problems that are multistep, need formulation, have more than one solution, and involve optimization. The findings suggest that current educational practices do not prepare students well for solving these types of problems. College and middle school students in these studies were good calculators but, from our perspective, were relatively weak problem solvers. They frequently omitted important plan elements, failed to test their solutions against constraints, and usually considered only one plan. Furthermore, even the dyads whose protocols we characterized as locally coherent (as defined by statement pairs) did not verbalize an overall plan for solving the challenge; that is, they did not first articulate an ordered set of goals and subsequently evaluate their progress relative to these goals. Students' paper artifacts also suggest a lack of planning. Students tended to write anywhere on the page and did not label their answers or otherwise document their work. As a consequence, students frequently lost track of what they had already done.

We argue that more materials that afford these kinds of complex problem-solving experiences are needed. The *Jasper* materials represent one set of curriculum resources, but there are other excellent examples of ways to enrich problem-solving activities in classrooms. We have in mind work by Resnick and Bill (e.g., Resnick, Bill, Lesgold, & Leer, 1991), Ball (1993), Cobb, Yackel, and Wood (e.g., Cobb, Wood, & Yackel, 1991; Cobb et al., 1992; T. Wood et al., 1992; T. Wood & Yackel, 1990; Yackel, Cobb, & Wood, 1991), and Lampert (1990). Recent work on project-based learning also holds promise (elsewhere we discuss effective design features of projects; see Barron et al., 1995; Barron et al., 1996).

Although this research was conducted in a laboratory setting, we believe that our results have implications for collaborative problem solving in classroom settings. Our findings indicate that, in the absence of instruction, some students will interact with their peers in ways that promote exploration of the problem space and its solution. Even among our sample of high-achieving mathematics students, however, there were individual differences in their patterns of interaction. We would expect similar variation among average- and low-achieving students and, consequently, that many students could benefit from intervention on effective collaborative communication.

What are the instructional supports in traditional classrooms that might help students acquire effective patterns of argumentation? We argue that current practices in many, if not most, mathematics classrooms are not supportive of the kind of peer interaction that was predictive of more successful problem solving in this research. (We are referring to classrooms where more traditional, direct instruction is the norm; we are not referrring to constructivist-oriented classrooms where conjecture and argumentation are fundamental to classroom discussions.) For example, the discourse that is the norm in traditional whole class discussions tends to focus on teachers telling or eliciting the telling of rote procedures for manipulating numbers in order to produce "correct answers." There is virtually no peer-to-peer interaction, and there is little emphasis on understanding the underlying mathematics concepts and procedures. These types of whole class discussions are poor models for how students might best interact with their peers in small groups. Furthermore, small group collaborative work in classrooms is often tutorial in nature; teachers pair students so that one student who has mastered a rote procedure can tell the other student the appropriate steps to follow to find "the" answer. Again, there is a disjuncture between the norms often operative in the tutorial situation and the norms that we think will be supportive of effective argumentation patterns. Note that our assumption is that effective argumentation needs to be established as a classroom norm. If there are norms in place in some activity structures, such as whole class discussions, which are inconsistent with patterns of effective argumentation, it will be difficult to establish other norms in small collaborative groups.

We assume that the norms for discourse in whole class discussions can and should be similar to the discourse patterns of the effective dyads in Experiment 2. We further assume that classroom norms that best support effective collaboration will be consistent with constructivist-oriented pedagogy.[2] We speculate that, to promote effective interactions, it will be important to emphasize deep understanding of concepts and procedures. When students are engaged in the process of sense making, it is functional for students to share their conceptions and to react to the conceptions of other students (Goldman, 1997; Secules et al., 1997; Vye et al., in press). In contrast, discussions about rote procedures or algorithms will, almost by definition, focus on the "right way" to do it rather than engage reasoning and problem-solving processes. We also assume that teachers will need to set expectations and model certain behaviors for their students in both whole class and small group settings; for example, to emphasize that students listen and react to the thoughts of their peers, explain their answers and thinking, and conjecture about concepts or strategies.[3] We assume that teachers will need to remove themselves from the role of intellectual authority of the classroom. They will need to scaffold peer interactions in such a way that students come to appreciate that they can resolve differences of opinions and draw appropriate conclusions without appealing to the teacher for the "right answer."

[2]We are grateful to one anonymous reviewer for pointing out the importance of engaging in conjecture that, in this context, we take to mean a willingness to articulate one's perspective in the face of some uncertainty and to reason about ways to validate or evaluate this perspective. We agree that effective argumentation does not simply involve "telling what you know."

[3]As one anonymous reviewer of this article pointed out, the whole class discourse in Deborah Ball's (e.g., Ball & Rundquist, 1993) mathematics classroom is exactly the sort that occured with the dyads in this study.

ACKNOWLEDGMENTS

The research reported in this article was supported in part by James S. McDonnell Foundation Grant 91–6 and by National Science Foundation Grant MDR–9050191. Any opinions expressed herein are our own and do not necessarily reflect those of the granting agencies. Members of the Cognition and Technology Group at Vanderbilt who contributed to this article are John Bransford, Cynthia Mayfield-Stewart, Allison Moore, James S. Pellegrino, and Kirsten Rewey. We would also like to acknowledge Mary Means for her invaluable contributions to the research.

REFERENCES

American Association for the Advancement of Science. (1989). *Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology.* Washington, DC: Author.

Angell, R. B. (1964). *Reasoning and logic.* New York: Appleton–Century–Crofts.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal, 93,* 373–397.

Ball, D. L., & Rundquist, S. S. (1993). Collaboration as a context for joining teacher learning with learning about teaching. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 13–42). San Francisco: Jossey-Bass.

Barron, B. (1991). *Collaborative problem solving: Is team performance greater than what is expected from the most competent member?* Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Barron, B., Schwartz, D., Vye, N. J., Moore, A., Petrosino, A., Zech, L., Bransford, J. D., & Cognition and Technology Group at Vanderbilt. (1996). *Doing with understanding: Lessons from research on problem- and project-based learning.* Unpublished manuscript, Vanderbilt University, Nashville, TN.

Barron, B., Vye, N. J., Zech, L., Schwartz, D., Bransford, J. D., Goldman, S. R., Pellegrino, J., Morris, J., Garrison, S., & Kantor, R. (1995). Creating contexts for community-based problem solving: The Jasper challenge series. In C. N. Hedley, P. Antonacci, & M. Rabinowitz (Eds.), *Thinking and literacy: The mind at work* (pp. 47–71). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–41.

Chi, M. T. H., Glaser, R., & Farr, M. (1991). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cobb, P., Wood, T., & Yackel, E. (1991). Analogies from the philosophy and sociology of science for understanding classroom life. *Science Education, 75,* 23–44.

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal, 29,* 573–604.

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(6), 2–10.

Cognition and Technology Group at Vanderbilt. (1991). Technology and the design of generative learning environments, *Educational Technology, 31*(5), 34–40.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper experiment: An exploration of issues in learning and instructional design. *Educational Technology Research and Development, 40,* 65–80.

Cognition and Technology Group at Vanderbilt. (1993). The Jasper series: Theoretical foundations and data on problem solving and transfer. In L. A. Penner, G. M. Batsche, H. M. Knoff, & D. L. Nelson (Eds.), *The challenges in mathematics and science education: Psychology's response* (pp. 113–152). Washington, DC: American Psychological Association.

Cognition and Technology Group at Vanderbilt. (1994). From visual word problems to learning communities: Changing conceptions of cognitive research. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 157–200). Cambridge, MA: MIT Press/Bradford Books.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper project: Lessons in curriculum, instruction, assessment, and professional development.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dewey, S. (1933). *How we think: Restatement of the relation of reflective thinking to the educative process.* Boston: Heath.

Fay, A. L., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development, 67,* 689–716.

Glaser, R. (1991). Intelligence as an expression of acquired knowledge. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 47–56). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Glaser, R. (1994). Learning theory and instruction. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Eds.), *International perspectives on psychological science: Vol. 2. The state of the art* (pp. 341–357). Hove, UK: Lawrence Erlbaum Associates, Inc.

Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes, 23,* 357–398.

Goldman, S. R., & Cognition and Technology Group at Vanderbilt. (1991, August). *Meaningful learning environments for mathematical problem solving: The Jasper problem solving series.* Paper presented at the European Conference for Research on Learning and Instruction, Turku, Finland.

Goldman, S. R., Cosden, M. A., & Hine, M. S. (1992). Working alone and working together: Individual differences in the effects of collaboration on learning handicapped students' writing. *Learning and Individual Differences, 4,* 369–393.

Greeno, J. G. (1986). Collaborative teaching and making sense of symbols: Comment on Lampert's "Knowing, doing, and teaching multiplication." *Cognition and Instruction, 3,* 343–348.

Heaton, R. M., & Lampert, M. (1993). Learning to hear voices: Inventing a new pedagogy of teacher education. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 43–83). San Francisco: Jossey-Bass.

Hine, M. S., Goldman, S. R., & Cosden, M. A. (1990). Error monitoring by learning-handicapped students engaged in collaborative microcomputer-based writing. *The Journal of Special Education, 23,* 407–422.

Howe, C., Tolmie, A., Greer, K., & Mackenzie, M. (1995). Peer collaboration and conceptual growth in physics: Task influences on children's understanding of heating and cooling. *Cognition and Instruction, 13,* 483–504.

Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27,* 65–90.

Lamon, M., Secules, T. J., Petrosino, A. J., Hackett, R., Bransford, J. D., & Goldman, S. R. (1996). Schools for thought: Overview of the project and lessons learned from one of the sites. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 243–288). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lampert, M. (1986). Knowing, doing, and teaching multiplication. *Cognition and Instruction, 3,* 305–342.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27,* 29–63.

Mason, J., Burton, L., & Stacey, K. (1982). *Thinking mathematically.* London: Addison-Wesley.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Engelwood Cliffs, NJ: Prentice Hall.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Peterson, P. L., Carpenter, T., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *Journal of Educational Psychology, 81,* 558–569.

Peterson, P. L., Fennema, E., & Carpenter, T. (1991). Using children's mathematical knowledge. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), *Teaching advanced skills to at-risk students* (pp. 68–111). San Francisco: Jossey-Bass.

Peterson, P. L., Fennema, E., Carpenter, T., & Loef, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and Instruction, 6,* 1–40.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44,* 162–169.

Resnick, L. B., Bill, V. L., Lesgold, S. B., & Leer, M. N. (1991). Thinking in arithmetic class. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), *Teaching advanced skills to at-risk students* (pp. 27–53). San Francisco: Jossey-Bass.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13,* 89–99.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences, 1,* 37–68.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic.

Schoenfeld, A. H. (Ed.). (1987). *Cognitive science and mathematics education.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Secules, T., Cottom, C., Bray, M., Miller, L., & The Schools for Thought Collaborative. (1997). Creating schools for thought. *Educational Leadership, 54*(6), 56–60.

Van Haneghan, J. P., Barron, L., Young, M. F., Williams, S. M., Vye, N. J., & Bransford, J. D. (1992). The Jasper series: An experiment with new ways to enhance mathematical thinking. In D. F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 15–38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

VanLehn, K., & Brown, J. S. (1980). Planning nets: A representation for formalizing analogies and semantic models for procedural skills. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Vol. 2. Cognitive process analyses of learning and problem solving* (pp. 95–137). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Voss, J. F. (1991). Informal reasoning and international relations. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 37–58). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Voss, J. F., Blais, J., Means, M. L., Greene, T. R., & Ahwesh, E. (1986). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. *Cognition and Instruction, 3,* 269–302.

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction, 16,* 337–350.

Vye, N. J., Schwartz, D. L., Bransford, J. D., Barron, B. J., Zech, L. K., & Cognition and Technology Group at Vanderbilt. (in press). SMART environments that support monitoring, reflection, and revision. In D. J. Hacker, J. Dunlosky, & A. C. Graessar (Eds.), *Metacognition in educational theory and practice.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press. (Original work published 1934)

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21–39.

Webb, N. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22,* 366–389.

Webb, N., Ender, P., & Lewis, S. (1986). Problem solving strategies and group processes in small group learning computer programming. *American Educational Research Journal, 23,* 245–261.

Webb, N., & Palincsar, A. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). New York: MacMillan.

Whitehead, A. N. (1929). *The aims of education.* New York: MacMillan.

Wood, D., Wood, H., Ainsworth, S., & O'Malley, C. (1995). On becoming a tutor: Toward an ontogenetic model. *Cognition and Instruction, 13,* 565–582.

Wood, T., Cobb, P., & Yackel, E. (1992). Change in learning mathematics: Change in teaching mathematics. In H. H. Marchall (Ed.), *Redefining student learning: Roots of educational change* (pp. 177–205). Norwood, NJ: Ablex.

Wood, T., & Yackel, E. (1990). The development of collaborative dialogue within small group interactions. In L. P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education: International perspectives* (pp. 244–252). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education, 22,* 390–408.

APPENDIX

Portion of the Expanded Goal Structure Used in Goals, Reasoning, and Arguments Analysis

- Evaluate acceptable plans with respect to profit, risk, and conservation
  - Estimate profit
    - Fire department
    - School hose
    - Pool truck
    - Pool truck and school hose
  - Determine risk to fill pool
    - Risk for fire department option
  - Determine conservation problems
    - Problem of letting water run out on the ground
  - Estimate income by extrapolation ($270.00)
    - Determine no. of students at school (380 – 20)
    - Determine extrapolation factor (360/60)
    - Determine ticket price that yields highest income ($1,000)
      - Find sample income at each price
        - Sample income at $.50 ($29.00)
        - Sample income at $1.00 ($45.00)
        - Sample income at $1.50 (24 × $1.50)
        - Sample income at $2.00 (8 × $2.00)
        - Find no. of students at each price
          - No. of students at $.50 (58)
          - No. of students at $1.00 (45)
          - No. of students at $1.50 (16 + 8)
          - No. of students at $2.00 (8)