

The Perils of Averaging Data Over Strategies: An Example From Children's Addition

Robert S. Siegler
Carnegie-Mellon University

SUMMARY

If a person uses different strategies on different trials, averaging data generated by those strategies can distort conclusions about numerous aspects of performance. The present study illustrates both the dangers of averaging data generated by different strategies and the gains that can be realized by examining performance generated by each strategy separately.

The context for investigating these issues was young elementary-school children's addition. Previous models have depicted young children as always solving simple addition problems by using the *min* strategy, in which they count up from the larger number being added. For example, they would solve $3 + 6$ by starting at 6 and counting from there to 9. The conclusion that children consistently use this approach to solve simple addition problems has been based primarily on the results of chronometric analyses that have indicated that both individuals and groups of children show the solution time pattern predicted by the min model.

The present study involved examining children's verbal reports of their strategy on each problem as well as their solution-time patterns. When solution times on all trials were analyzed together, as in earlier studies, the results were entirely consistent with the view that children always use the min strategy. However, the verbal reports suggested a quite different picture. The min strategy was but one of five approaches that children reported using. This use of diverse strategies characterized individual as well as group performance; most children reported using at least three approaches. Neither the min strategy nor any other approach was used on as many as 40% of trials.

Considerable converging evidence supported the validity of the children's verbal reports. Most important, on trials where they reported using the min strategy, the min model was an even better predictor of solution times than in past studies or in the present data set as a whole. In contrast, on trials in which they reported using one of the other strategies, the min model was never a good predictor of performance, either in absolute terms or relative to other predictors.

Three factors that can lead to incorrect conclusions about data averaged over strategies were identified: relative frequency of each strategy, relative variability of performance generated by each strategy, and independent-dependent variable relations across and within strategies. The influence of these factors was illustrated both with regard to the present data and, for the more general case, through analyses of synthetic data in which each factor's contribution could be independently examined.

There is no reason to think that variability of strategy use within a single person is limited to children or to arithmetic. Previous reports suggest that it is characteristic of adults as well as children and of tasks as diverse as spelling words, telling time, mentally rotating objects, and solving series completion problems. Among the other issues discussed are implications of the findings for when verbal self-reports are most useful, how people choose among alternative strategies, and how strategy-choice procedures develop.

People often know many strategies for solving a single problem. Averaging data generated by use of these different strategies carries the same risks as averaging data generated by different individuals. Just as data aggregated over people may not accurately reflect the behavior of any person (Estes, 1956), so data aggregated over strategies may not accurately reflect the

characteristics of any strategy. The undesirability of averaging over strategies has been noted previously (Newell, 1973), but the practice continues to be extremely common.

The goals of the present article are to describe in detail the problems that can arise from aggregating data produced by different strategies and to illustrate the types of gains that can

follow from separately examining performance generated by each strategy. These goals are pursued in the context of children's addition. In particular, I review current theories of young elementary-school children's addition, identify potential distortions in these theories introduced by averaging performance generated by different strategies, and consider changes that would be likely to emerge from analyzing separately data produced by different strategies. Then I report an experiment that directly contrasts the results that emerge from analyzing data averaged over strategies with those that emerge from analyzing separately data produced by different strategies. The analysis and experiment together illustrate not only that large differences emerge when data generated by each strategy are examined separately, but also the sources of particular incorrect conclusions within previous models. Finally, a number of general issues are considered, among them the types of tasks on which people are likely to use multiple strategies and how use of multiple strategies can be assessed.

Children's addition offers a particularly clear case for illustrating the effects of averaging over strategies. It is an area in which use of diverse strategies has often been reported (e.g., Carpenter & Moser, 1982; Fuson, 1982; Houlihan & Ginsburg, 1981). It also is an area in which strategy use on a given trial can be assessed. Both observation of overt behavior and asking children immediately after each trial how they solved the problem have proved useful in assessing strategy use (Hamann & Ashcraft, 1985; Houlihan & Ginsburg, 1981; Ilg & Ames, 1951; Siegler & Shrager, 1984; Svenson, Hedenborg, & Lingman, 1976). Another advantage of addition as an area for studying the impact of averaging over strategies is that previous analyses of averaged data have seemed so successful. They have provided the basis for several well-defined and influential models that make unambiguous, testable predictions about performance (Ashcraft, 1982; in press; Groen & Parkman, 1972). Finally, the unique precision of numbers as stimuli and the unambiguous predictions of the models that have been proposed allow clear contrasts to be drawn between conclusions that follow from averaging data over strategies and those that follow from examining separately data produced by each strategy.

Children's addition is an important topic in its own right. Children throughout the world learn to add, they spend a great deal of time mastering the skill, and their more advanced understanding of arithmetic and algebra directly depends on it. At the same time, it is essential to recognize that the central issue in this article, the dangers of averaging performance produced by

different strategies, is limited neither to children nor to arithmetic. Individuals have been found to use multiple strategies on tasks as diverse as spelling, series completion, question answering, and missionaries-and-cannibals problems (LeFevre & Bisanz, 1986; Reder, 1987; Siegler, 1986; Simon & Reed, 1976). Whenever a person uses multiple strategies, analyses that fail to separate data produced by the different strategies are likely to encounter the perils described in the present analysis of children's addition. The concluding section of the article explicitly examines the types of tasks on which such problems are most likely to arise and several means for coping with the problems.

Current Approaches to Children's Addition

Previous research suggests two quite dissimilar views of young elementary-school children's addition. The two views correspond to two methodological approaches that have been used to study the subject. One approach has emphasized the construction of chronometric models based on patterns of mean or median solution times on different problems (Ashcraft, 1982; in press; Groen & Parkman, 1972; Kaye, 1986; Svenson, 1975). A key assumption of these models is that the best predictor of solution times can be used to infer the strategy that generated the times. For example, Groen and Parkman (1972) observed that the size of the *smaller addend* (the smaller of the two numbers being added) was the best predictor of first graders' solution times; it accounted for roughly 70% of the variance in time to solve different problems. This led them to postulate that children of this age consistently use the *min strategy* to solve simple addition problems. This min strategy involves counting up from the larger addend the number of times indicated by the smaller addend. For example, a child using the min strategy would solve $7 + 2$ by starting at 7 and counting upward 2 counts (i.e., counting "8, 9"). Groen and Parkman hypothesized that the only source of variation in solution times for different problems was the number of counts upward from the larger addend needed to solve the problem. Thus, $4 + 3$, $3 + 7$, and $9 + 3$ all would produce the same solution times, because all require 3 counts upward from the larger number.

The findings on which the min model was based have been replicated by numerous other investigators using chronometric methods (e.g., Ashcraft, 1982; Groen & Resnick, 1977; Kaye, Hall, Post, & Dineen, 1986; Svenson, 1975). They also have been extended to older children. In one particularly influential study, Ashcraft (1982) found that the size of the smaller addend was the best predictor of first graders' solution times, that the size of the smaller addend and the size of the sum squared were equally good predictors of third graders' times, and that the size of the sum squared was the best predictor of the times of children beyond third grade. This led him to conclude that first graders consistently used the min strategy, that fourth graders and older children consistently used retrieval, and that third graders sometimes used the min strategy and sometimes retrieval. Recent revisions of Ashcraft's model have recognized that first graders also at times use retrieval (Ashcraft, in press; Hamann & Ashcraft, 1985). However, the min strategy has continued to be viewed as first graders' dominant procedure, as evidenced by the fact that within the Ashcraft (in press) computer

This research was supported in part by National Institutes of Health Grant HD-19011 and in part by National Institute of Education Grant 83-0050.

Thanks are due to Miss Hawkins, Miss Perry, and the teachers and students of University Park Elementary of Monroeville, Pennsylvania, who made the research possible. Thanks also are due to Mark Ashcraft, Sam Glucksberg, Eric Jenkins, David Klahr, Steve Reed, and James Staszewski, who offered valuable comments on previous versions of the manuscript.

Correspondence concerning this article should be addressed to Robert S. Siegler, Psychology Department, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

simulation of first graders' performance, the min strategy is always used on all but the simplest problems.

A variety of findings support the chronometric models' emphasis on young elementary-school children's use of the min strategy. As noted above, the size of the smaller addend is the single best predictor of the children's solution times. It is a good predictor in absolute as well as relative terms, accounting for between 60% and 77% of the variance in solution times in many studies. These studies have included quite diverse problem sets. They also have included children in special classes for poor mathematics achievement as well as children in standard classes, and children in Europe as well as in North America (Svenson & Broquist, 1975). Furthermore, the min model fits individual children's solution time patterns as well as group averages (Kaye et al., 1986; Groen & Resnick, 1977; Svenson et al., 1976).

The other main methodological approach has emphasized interviews in which children are asked to explain how they solved each problem. This method has led to an emphasis on the diversity of children's strategies. Individuals as well as groups of children report using multiple approaches. Typically, their reports have been classified into 5 to 10 strategies. Three of the best documented are the min strategy, retrieval, and the counting-all strategy, in which children count from one the number of times indicated by the sum (Baroody & Ginsburg, 1986; Carpenter & Moser, 1982; Fuson, 1982; Houlihan & Ginsburg, 1981; Ilg & Ames, 1951; Secada, Fuson, & Hall, 1983).

The interview studies provide additional evidence that young children sometimes use the min strategy, but also suggest that they do not always do so. For example, Houlihan and Ginsburg (1981) and Carpenter and Moser (1982) found that on single-digit addition problems, first and second graders reported using the min strategy on fewer than one half of trials.

Reconciling Findings From the Two Approaches

Potential Biasing Factors in Verbal Reports

How can the depictions of the chronometric and the interview studies be reconciled? One possibility is that children's verbal reports are inaccurate. The reports, although given soon after the problem is completed, are retrospective. Even adults' retrospective reports of problem solving activities are often misleading (Ericsson & Simon, 1984; Nisbett & Wilson, 1977); young children might be expected to give even less accurate verbal protocols. The verbal reports have not been related to detailed analyses of solution time and error patterns produced by each individual strategy; thus, convergent validation for the reports is lacking. In sum, the lack of correspondence between the chronometric models and the verbal reports may stem from the verbal reports inaccurately reflecting children's solution strategies.

Potential Biasing Factors in Chronometric Analyses

Another possibility is that the verbal reports are accurate and the chronometric analyses misleading. Three factors that may conspire to bias the chronometric analyses are relative frequency of the strategies, relative variability of dependent mea-

sure scores produced by them, and relations between independent and dependent variable values across and within strategies. Existing empirical data suggest that each of these factors may have helped create the impression that young children use the min strategy more than they actually do. The analyses of hypothetical data in the Appendix show exactly how they could create this impression.

Relative frequency of use of each strategy is the most obvious of the factors. The more frequently a strategy is used, the greater its impact in influencing the results of a data analysis that combines data generated by different strategies. Although young elementary-school children may not use the min strategy on a majority of trials, they still may use it more often than any other single approach. The children in both Houlihan and Ginsburg's (1981) and Carpenter and Moser's (1982) interview studies reported using the min procedure on a plurality, but not a majority, of trials.

Next consider the effects of strategies producing unequal variance on the dependent measure. If people use two strategies equally often on each problem, but more variance on the dependent variable is contributed by trials on which one of the strategies is used, that strategy will have the larger influence on the distribution of averaged data. In fact, as illustrated in the Appendix, a less frequently used strategy can have a greater effect on the overall data pattern than a more frequently used one, if the differences in variability on the dependent measure are sufficiently great. In the present context, the effects of differential variability may serve to exaggerate the apparent importance of the min strategy relative to retrieval. Consistent with this possibility, Hamann and Ashcraft (1985) found that solution times of first graders, who presumably use the min strategy relatively often, were far more variable over problems than those of seventh and tenth graders, who presumably retrieve answers to virtually all simple addition problems.

A third factor is relations among independent and dependent variables associated with each strategy. When data have been averaged over strategies, the relation between each predictor and the data reflects not only how well the predictor fits data produced by the strategy that is theoretically associated with it, but also how well it predicts data generated by other strategies. For example, how well the smaller addend predicts solution times on addition problems reflects not only how well it predicts times on min strategy trials but also how well it predicts solution times generated by other strategies such as retrieval. It seemed likely that the size of the smaller addend would be at least moderately successful in predicting performance produced by other strategies. Even with college students, who presumably consistently use retrieval on single-digit addition problems, the size of the smaller addend correlates significantly and positively with solution times (Ashcraft & Battaglia, 1978). Such correlations between the smaller addend and performance generated by approaches other than the min strategy would contribute to the impression that the min strategy was used more often than it actually was.

Implications for Current Models

The results of the interview studies, together with the potential sources of bias in the chronometric analyses, render it likely

that young children use diverse strategies to add. If this is so, what revisions in current models of children's addition will be needed? Parameter estimates, predictors of problem difficulty, and conditions under which each strategy is used are three likely areas of change.

Parameter Estimates

The regression equation used to express the min model is a simple linear equation with one free parameter, $y = a + bx$. In this equation, y is the total time to solve the problem, a is the intercept, b is the slope, and x is the size of the smaller addend. Within the model, a , the intercept, is viewed as reflecting the amount of time needed for performing all processes other than counting, whereas b , the regression coefficient of the smaller addend, is interpreted as the amount of time needed for each count when executing the min strategy. On the basis of values of b , investigators have estimated first and second graders' time per count as between 400 ms and 1 s, with the particular value depending on the regression coefficient in the particular experiment (Ashcraft, 1982; Ashcraft, Fierman, & Bartolotta, 1984; Groen & Parkman, 1972; Kaye et al., 1986; Suppes & Groen, 1967).

To the extent that the mean solution time for each problem reflects averaging over strategies, however, all of the estimates of the time per count when using the min strategy may be too low. In particular, the regression coefficient may include counting trials, where the slope is greater than the average, and retrieval trials, where it is smaller. If this is true, the estimated time per count when using the min strategy will need to be revised upward.

Predictors of Problem Difficulty

The best predictors of problem difficulty may also appear quite different when data produced by different strategies are examined separately. On trials in which the min strategy is used, the size of the smaller addend may be an even better predictor of the relative length of solution times on different problems than it has appeared to be in the past. Averaging data that were not produced by the min strategy with data that were produced in this way could dilute the relation between the size of the smaller addend and the solution time.

In contrast, solution times produced by other strategies would be expected to be best predicted by variables other than the size of the smaller addend. When children use the counting-all strategy, their solution times would be expected to be a function of the total number of counts, that is, a function of the sum. When children use retrieval, their solution times should be a function either of the size of the sum, like those of preschoolers (Siegler & Shrager, 1984), or of the size of the squared sum, like those of older children and adults (Ashcraft & Battaglia, 1978; Ashcraft, 1982).

A parallel pattern is expected on error percentages for each problem. The size of the smaller addend should be the best predictor of percentage of errors on min strategy trials, but not on trials where children use other strategies.

Conditions of Strategy Use

The third type of change involves abandoning an old question and raising a new one. The question "What is the strategy that young children use to add" makes little sense if even an individual child uses multiple strategies. Instead, it becomes more reasonable to ask "Under what conditions do children most often use each of their strategies?"

One way to address this question would be to examine how well various characteristics of problems predict frequency of use of each strategy, in much the same way that such characteristics are used to predict lengths of solution times and percentages of errors on each problem. Predicting when each strategy is most often used may be a more complex task, however. The difficulty is that the frequency with which a strategy is used on a problem is likely to reflect not only how effectively that strategy can be executed on the problem but also how effectively other strategies can be.

Consider a hypothetical example of use of the min strategy on two problems: $2 + 1$ and $7 + 8$. In absolute terms, young elementary-school students might use the min strategy on $2 + 1$ and on $7 + 8$ on similar, low percentages of trials. The reasons could be quite different, though. They might rarely use the min strategy on $2 + 1$ because they could accurately retrieve the answer to it. They might rarely use the min strategy on $7 + 8$ because there was little likelihood that they could execute the strategy accurately. These similar outcomes for different reasons would make it very difficult to isolate factors that influence children's decisions to use a strategy on the basis of absolute frequency of use of the strategy.

What would be a more useful dependent measure for determining when children use each strategy? Because the availability of multiple strategies is likely to affect the conditions under which each strategy is used, it may be useful to examine the form that the interaction among strategies is likely to take. Research on preschoolers' addition and subtraction (Siegler, 1986; Siegler & Shrager, 1984) suggests that the influence of faster and slower strategies on each others' use is asymmetric. If faster strategies can be executed accurately on a problem, they will be adopted regardless of the accuracy of slower strategies on the problem. In contrast, slower strategies will be used only if faster ones cannot be executed accurately.

This analysis suggests a dependent measure that may be a useful supplement to absolute frequency of strategy use for understanding the factors that influence children's strategy choices. This dependent measure is a conditional probability: probability of use of a strategy on a problem, given that a faster strategy was not used on the problem.

The problems $2 + 1$ and $7 + 8$ can again be used to illustrate. For simplicity, assume that children used three strategies—retrieval, the min strategy, and counting all—and that retrieval was fastest, followed by the min strategy and then by counting all. Assume further that on $2 + 1$, children used retrieval on 70% of trials, the min strategy on 20%, and counting all on 10%, whereas on $7 + 8$ they used retrieval on 10%, the min strategy on 20%, and counting-all on 70%. On $2 + 1$, the conditional probability of using the min strategy would be the 20% absolute probability of using the min strategy divided by the 30% proba-

bility of not using the faster retrieval strategy, that is .67. On $7 + 8$, the conditional probability of using the min strategy would be the 20% absolute probability of using the min strategy divided by the 90% probability of not using the faster retrieval strategy, that is .22. The conditional probabilities would show that if children did not state a retrieved answer, they were more likely to use the min strategy on $2 + 1$, on which it was easy to execute correctly, than on $7 + 8$, on which correct execution was more difficult. In contrast, the absolute probabilities would have indicated only that both strategies were used on 20% of trials on both problems. Over a set of problems, the pattern of conditional probabilities might indicate the specific ways in which each strategy's relative accuracy on different problems influenced its likelihood of being adopted. Thus, it seemed worth computing such conditional probabilities and examining whether they proved useful for understanding the factors influencing when each strategy was adopted.

How would the accuracy of each strategy be expected to influence its conditional probability of use? Consider the predictions for retrieval, counting all, and the min strategy. Retrieval was expected to have the highest conditional probability of use on problems with small sums and squared sums, because early in the learning process, these are the problems on which it most often produces accurate performance (Siegler & Shrager, 1984). The conditional probability of using the counting-all strategy also was expected to be highest on problems with small sums and squared sums, for the same reason. The prediction for the min strategy was more complex, because it depends on whether accuracy of the strategy is defined in absolute terms or relative to other strategies. Problems with a small minimum addend would be the ones on which the min strategy would yield its most accurate performance in absolute terms. However, problems with large differences between the sizes of the addends would produce the greatest advantage in the accuracy of the min strategy relative to that of other strategies. The smaller addend on such problems is by definition small relative to the sum and the sum squared. This would make accurate execution of the min strategy easy relative to accurate execution of retrieval or counting all. Thus, conditional probability of use of the min strategy was expected to be highest on problems with small minimum addends or large differences between addends.

Experimental Predictions

To summarize, it was expected that separately analyzing data generated by different strategies would result both in revisions of previous conclusions about children's addition and in new findings on issues that have not been addressed previously. Among the anticipated revisions were that the min strategy would be used much less often than implied by current chronometric models, that the time per count using the min strategy would be higher than previously estimated, and that size of the smaller addend would be the best predictor of solution times and errors only on the subset of trials on which the min strategy was used. Among the new findings that were anticipated were that children would use each addition strategy most often when the strategy would lead to favorable combinations of speed and accuracy relative to those yielded by other strategies, and that

conditionalizing probability of strategy use on not using faster strategies would prove useful for understanding exactly how a strategy's accuracy on different problems influenced when it was used. More generally, the central purpose of the experiment was to demonstrate how verbal protocol, solution time, and error data can be used together to identify each person's problem solving strategy on each trial, and to illustrate the value of such specific assessments for understanding when strategies are used and how they are chosen.

Method

Participants

The children were 22 kindergarteners, 28 first graders, and 18 second graders. The kindergarteners were students at a university-run preschool; the first and second graders attended an upper-middle-class suburban public school. The kindergarteners included 12 boys and 10 girls, the first graders, 16 boys and 12 girls, and the second graders, 9 boys and 9 girls. The mean ages were 69 months, 80 months, and 92 months for the kindergarteners, first graders, and second graders, respectively. The kindergarteners had received some exposure to single-digit arithmetic problems with sums less than 10 as part of their daily routine for the three months prior to testing. The first graders had substantial experience with problems whose sums were 10 or less and had received a small amount of instruction both on problems with sums between 10 and 18 and on problems with larger sums that did not require carrying. The second graders had received substantial amounts of instruction on both single- and multiple-digit problems, including ones that require carrying. A 33-year-old female research assistant was the experimenter.

Problems

Children were presented with 45 problems, 9 each of five types. As is shown in Table 1, the problem types varied along a number of dimensions that might influence the frequency, speed, and accuracy of each strategy: minimum addend, sum, sum squared, addend order, and difference between the smaller and larger addend. The problems were chosen to minimize correlations among these variables, so as to maximize discrimination among alternative interpretations of the data. With the exception of the unavoidably high correlation between the sizes of the sum and the squared sum ($r = .97$), the correlations among the 10 possible pairs of the 5 variables were moderate (from $r = -.33$ to $r = .61$).

The Table 1 problems allow clear discrimination among variables that might influence when each strategy is adopted. This can be illustrated with regard to the issue of when children use the min strategy. If they use the min strategy most often when all features of the problem would seem to favor it (a small minimum addend, which would make the strategy easy to execute; a large difference between addends, which would maximize the gain from counting-on from the larger addend relative to counting from one; and an addend order of the larger addend first, so that children could execute the strategy without transposing the addends), they would use the min strategy most often on Problem Type 1, which includes all of these features. Alternatively, if they used the min strategy whenever either ease of execution or a large gain relative to counting from 1 favored it, they would adopt the min strategy most often on Problem Types 1, 2, 4, and 5. Two other plausible alternatives were that they would use it most often whenever the minimum addend was small, which would lead to most frequent use on Problem Types 1, 2, and 5, or whenever the difference between addends was large, which would lead to most frequent use on Problem Types 1, 2, and 4.

Table 1
Types of Problems Presented to Children

Dimension	Problem type				
	1	2	3	4	5
Problems	(12-14) + (1-3)	(1-3) + (12-14)	(5-7) + (8-10)	(15-17) + (4-6)	(4-6) + (1-3)
Sum	13-17	13-17	13-17	19-23	5-9
Smaller addend	1-3	1-3	5-7	4-6	1-3
Difference	9-13	9-13	1-5	9-13	1-5
First addend	larger	smaller	smaller	larger	larger
Sum squared	169-289	169-289	169-289	361-529	25-81

Note. For each problem type, the items were the nine factorially possible combinations of the first and second addends indicated in the Problems row of the table.

Procedure

Each child was brought individually from the classroom to a vacant room within the school. The child was seated at a table directly across from the experimenter. Before the first session, the child was told,

We are going to do some addition problems today. I'll read a problem to you, and when you have an answer, tell me what it is. You can do anything you want to get the right answer. You can count or use your fingers or do whatever you want to do. It doesn't matter how you get the right answer as long as you try the best that you can.

After being asked the first problem in the first session, the child was told, "We're interested in knowing how children your age figure out the answers to these problems. Tell me, how did you figure out the answer to that problem?" The question "How did you figure out the answer to that problem" was repeated after each item, unless the child volunteered the information before being asked, which children usually did after a few items. If the child's description was unclear, the experimenter would ask follow-up questions. For example, if the child simply said, "I counted," the experimenter would ask, "What number did you begin your counting on?"

The 45 problems were presented to each child over 5 school days. A randomly chosen group of 9 problems (different for each child) was presented on each of the days. The days were as close to consecutive as possible, given constraints of the child's schedule. After the first day, the experimenter introduced the procedure by asking whether the child remembered the game they played earlier and indicating that they would play the same game again.

Each child's behavior was recorded with a Sony SLO-323 videocassette recorder and a Sony 3260 camera. Solution times were recorded with a Vicon X240 digitizer, which printed digital times across the bottom of the taped scene. The times were accurate to .1 s, which seemed to be a sufficient degree of accuracy for the present task, in which the median solution time was 4 s. The videotaped records proved to be a useful supplement to the children's explanations in cases where the explanations were unclear and in the few instances where the children's overt behavior contradicted their descriptions of what they had done.

The verbal explanations and videotaped record of ongoing performance allowed a good degree of interrater reliability in classifications of strategies. Two research assistants independently classified which strategy the child used on each trial after listening to and watching the videotape of it. The two raters agreed on 94% of their classifications. On the other 6%, they discussed their observations and reviewed the tape until they reached agreement. Agreement on the initial classifications of each of the strategies was at least 89%.

Results

Overview

Three types of data were of primary interest: data on strategy use, solution times, and errors. For each of these, it seems worthwhile to provide an overview before presenting more detailed analyses.

Children's reports indicated use of five approaches: retrieval, the min strategy, counting all, decomposition, and guessing. Retrieval, the min strategy, and counting all were described in the introduction. Decomposition involved transforming the original problem into two or more simpler problems. This usually meant separating an addend whose value exceeded 10 into 10 and a remaining portion, adding the remaining portion to the original smaller addend, and adding 10 to the sum. For example, a child asked to solve $12 + 2$ might say, "12 is 10 and 2; 2 and 2 is 4; 10 and 4 is 14; so 14." Occasionally, it involved referring to better known problems with similar addends (e.g., "7 plus 8 is like 7 plus 7 and 1 more, so 15"). The remaining classification, guessing, was used when children explained their answer by saying that they guessed or that they did not know the answer. Both min and counting all categories included trials in which children counted and put up fingers and trials in which they counted but did not put up fingers. Neither solution time nor error patterns differentiated the two cases, so they were grouped together.

Table 2 illustrates the percentage of use of each strategy by children in each grade. At all three grade levels, children used multiple strategies. There were also clear developmental trends toward increasing use of retrieval and decomposition, and decreasing use of counting all and guessing.

Children's multiple strategy use was apparent at the level of individual children as well as at the group level. Fully 99% of children reported using 2 or more strategies, and 62% reported using 3 or more. At least 68% of children at each age/grade level reported using both the min strategy and retrieval. The majority of kindergarteners also reported using counting all and guessing; the majority of first and second graders also reported using decomposition.

These individual data suggest that by first grade, most children knew all five strategies. Each strategy except decomposi-

Table 2
Percentage of Use of Each Strategy by Children of Each Age

Grade level	Strategy				
	Retrieval	Min	Decomposition	Count all	Guess or no response
Kindergarten	16	30	2	22	30
Grade 1	44	38	9	1	8
Grade 2	45	40	11	0	5
Overall	35	36	7	8	14

tion was cited by at least 68% of kindergarteners. The majority of first and second graders reported using decomposition, as well as retrieval and the min strategy. It seems likely that the other two approaches, counting all and guessing, were not forgotten between kindergarten and first grade. More likely, first and second graders could add faster and more accurately using other strategies, so they did not use the two, less efficient approaches that they also knew.

Perhaps the most striking finding in Table 2 is that the kindergarteners and first and second graders reported using the min strategy on only 36% of trials. In no grade did the percentage exceed 40%. These percentages differ dramatically from the chronometric models' implication that young elementary school children use the min strategy on all trials on most problems.

Measured either in terms of accuracy or speed, performance on these problems improved greatly over this age/grade range. As is shown in Table 3, the percentage of errors decreased from 51% to 5%; median solution time decreased from 6.0 s to 2.7 s.

This overview provides a context within which to consider more detailed analyses of solution time, error, and strategy-use patterns.

Solution Times

Analyses of times averaged over strategies. Traditionally, the central analysis used to infer children's addition strategy has been a multiple regression analysis of mean or median solution times on each problem. This type of analysis was also performed on the median solution times on each problem in the present experiment. The purpose was not to establish the strategy that children used, but to determine the comparability of the present data with those reported previously, and to establish a point of comparison for subsequent analyses in which data produced via different strategies were analyzed separately.

As in all of the regression analyses conducted in this study, nine predictor variables were included. Seven of these had been widely used in previous studies: the sizes of the first addend, the second addend, the larger addend, and the smaller addend; the difference between the addends; the sum; and the sum squared. The other two were included because of hypotheses specific to this study concerning the conditions under which different strategies would be used: whether one of the addends was larger than 10, and whether either the smaller addend was small (between 1 and 3) or the difference between addends was large

Table 3
Errors and Solution Times Generated by Each Strategy

Grade level	Strategy					All trials
	Retrieval	Min	Decomposition	Count all	Guess or no response	
Median solution times						
Kindergarten	3.9	6.0	6.9	15.2	4.2	6.0
Grade 1	2.1	6.9	4.1	16.3	7.4	3.8
Grade 2	1.8	3.9	3.2	—	3.7	2.7
<i>Mdn</i>	2.1	5.6	3.8	15.2	5.0	4.0
Percentage of errors						
Kindergarten	19	29	9	55	89	51
Grade 1	4	17	8	50	38	13
Grade 2	3	7	3	—	7	5
<i>M</i>	6	17	6	54	71	23

Note. The final row and the final column in each section of the table are not unweighted averages of the numbers in their column or row because the strategies were used on different numbers of trials and there were different numbers of children of each age. Instead, they are the overall mean or median for each age or for each strategy based on all responses.

(9 or more). Because of the children's high percentage of errors (23%), and because the errors were generated by the same strategies as the correct answers, both correct and incorrect trials were used in computing median solution times on each problem.

As in previous studies of young elementary-school children's addition, the best predictor of solution times averaged over strategies was the size of the smaller addend. The percentage of variance accounted for by this variable, 76%, was also comparable to that found in previous investigations (e.g., 76% by Groen & Parkman, 1972; 62% by Ashcraft, 1982; 77% and 62% by Ashcraft et al., 1984 on first graders' verification and production task performance, respectively). Also as in previous experiments, no other predictor accounted for the averaged data nearly as well as the size of the smaller addend. The next best predictor, the size of the squared sum, accounted for only 41%.

It has become standard practice in addition experiments where the size of the smaller addend is the best predictor of solution times to interpret the regression coefficient of the smaller addend as the time needed for each count within the min strategy. Interpreting the coefficient in this way suggested a counting rate of 927 ms/count ($SE = 78$ ms). This estimate was somewhat higher than the 660 ms/count that results from averaging the estimates of the five previous studies of children this age (cited in the introduction) that reported the counting rate, but was within the range of the reported times of 410–959 ms/count.

Analyses of times produced by each strategy. Separate multiple regression analyses were conducted for each strategy on the median solution times on each problem. Only problems on which the strategy under consideration was used at least three times were included in these analyses, so that no single outlying time could exert a large influence on the median time for a problem. This criterion allowed analyses of all 45 problems for the min strategy, 43 for retrieval, 34 for guessing, 30 for counting all, and 29 for decomposition.

As is shown in Table 4, separately analyzing the solution times produced by each strategy revealed a very different picture than the one yielded by analyzing solution times averaged over strategies. On trials in which children reported using the min strategy, the size of the smaller addend was by far the best predictor of solution times. It accounted for 86% of the variance. Thus, on those trials in which children reported using the min strategy, the size of the smaller addend accounted for considerably more variance than it did when solution times were averaged across all strategies in the present experiment, and considerably more variance than it had in any previously reported data set.

In contrast, the size of the smaller addend was not the best predictor of solution times when any of the other strategies were used. It was not even the second-best predictor in any of the analyses. On retrieval trials, it accounted for 21% of the variance; on decomposition trials, it accounted for 37%; and on both counting-all and guessing trials, it accounted for 12%. It did not add significant independent variance to the regression equations for any of these approaches. Thus, the size of the smaller addend was an excellent predictor of solution times on the 36% of trials in which children said they used the min strat-

Table 4
Best Predictor of Median Solution Time and Percentage of Errors on Each Problem

Strategy	Best predictor	R^2
Solution time		
All trials	Smaller addend	76
Retrieval	Sum	30
Min	Smaller addend	86
Decomposition	Sum squared	42
Count all	Sum	35
Guessing	No significant predictors	
% E		
All trials	Smaller addend	78
Retrieval	Sum squared	30
Min	Smaller addend	74
Decomposition	Sum	64
Count all	Sum	50
Guessing	No significant predictors	

egy, but only a fair predictor on the 64% of trials in which they said they used other approaches.

The estimated time per count (the regression coefficient of the smaller addend) also changed when data produced by different strategies were analyzed separately. On trials in which children said they used the min strategy, the regression coefficient indicated a counting time of 1.30 s/increment ($SE = .08$ s). This was 35% greater than the highest previously estimated increment per count, and 40% greater than the increment per count that would have been estimated from the median solution times on all trials in the present experiment.

Analyzing each strategy separately also indicated how the size of the smaller addend could be such a good predictor of the overall solution time pattern, despite the min strategy being used on only 36% of trials. The children used two strategies most often, the min strategy and retrieval. They used them on almost identical percentages of trials, 36% and 35%, respectively. The variability of median solution times on min strategy trials was much greater than the variability on retrieval trials ($SD = 2.72$ s vs. 0.73 s), which would tend to make the best predictor of performance on min strategy trials, the smaller addend, a better predictor of the averaged data than the best predictor on retrieval trials.

Independent-dependent variable relations within strategies and across strategies also contributed to the ability of the size of the smaller addend to so accurately predict the averaged solution times. Figure 1 illustrates these relations for the two frequently used approaches, the min strategy and retrieval. Note that the predictor associated with the min strategy, the size of the smaller addend, was a much better predictor of min strategy times, $r = .93$, than was the predictor associated with retrieval, the size of the sum squared, of retrieval times, $r = .54$. This superiority contributed to the size of the smaller addend being a better predictor of the averaged solution times than the size of the sum squared, despite the min strategy and retrieval being used equally often. The positive correlation between the size of

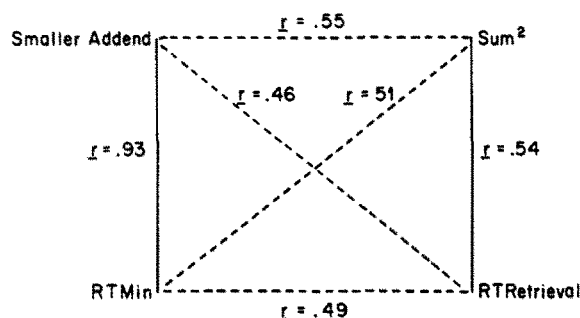


Figure 1. Zero-order correlations of independent and dependent variables for min strategy and retrieval trials. ("RTmin" and "RTretrieval" refer to solution times on min strategy and retrieval trials, respectively. Solid lines indicate correlations between predictor most closely associated with a strategy and median solution times on that strategy.)

the smaller addend and solution times on retrieval trials, $r = .46$, contributed to the size of the smaller addend being an excellent predictor of the averaged solution times in absolute as well as relative terms. Although it is not shown in the table, the size of the smaller addend also correlated positively with solution times generated by other strategies: $r = .61$ on decomposition trials, $r = .11$ on counting all trials, and $r = .34$ on guessing trials.

Figure 2 graphically illustrates how the size of the smaller addend could be such a good predictor of the averaged solution times. It shows that the size of the smaller addend was an excellent predictor of the highly variable times on min strategy trials, that it also was positively correlated with the much-less-variable times on retrieval trials, and that it could quite accurately pre-

dict the averaged data when times produced by the two strategies were combined.

In sum, the min strategy's relatively high frequency of use, the relatively high variability of solution times that it produced, the superior fit of the smaller addend to performance on min strategy trials, and the positive correlations between the smaller addend and performance produced by approaches other than the min strategy, together allowed the size of the smaller addend to be an excellent predictor of the averaged solution times, even though the min strategy was used on only 36% of trials.

Errors

Analyses of percentage of errors on each problem yielded similar results to the analyses of solution times (Table 4). Again, when data were averaged over strategies, the size of the smaller addend was the best predictor, accounting for 78% of the variance in the percentage of errors on each problem. Again, when data produced by each strategy were analyzed separately, the pattern changed dramatically. On trials in which children reported using the min strategy, the size of the smaller addend accounted for 74% of the variance in the percentage of errors on each problem. This was far more than the percentage of variance accounted for by the next best predictor of errors on these trials, 28% by the sum squared, though slightly less than the 78% of variance accounted for on all trials by the size of the smaller addend.

In contrast, on trials where children said they used other strategies, the size of the smaller addend was never either the best or the second best predictor. Nowhere other than the min strategy did it add significant independent variance to that which could be accounted for by other variables.

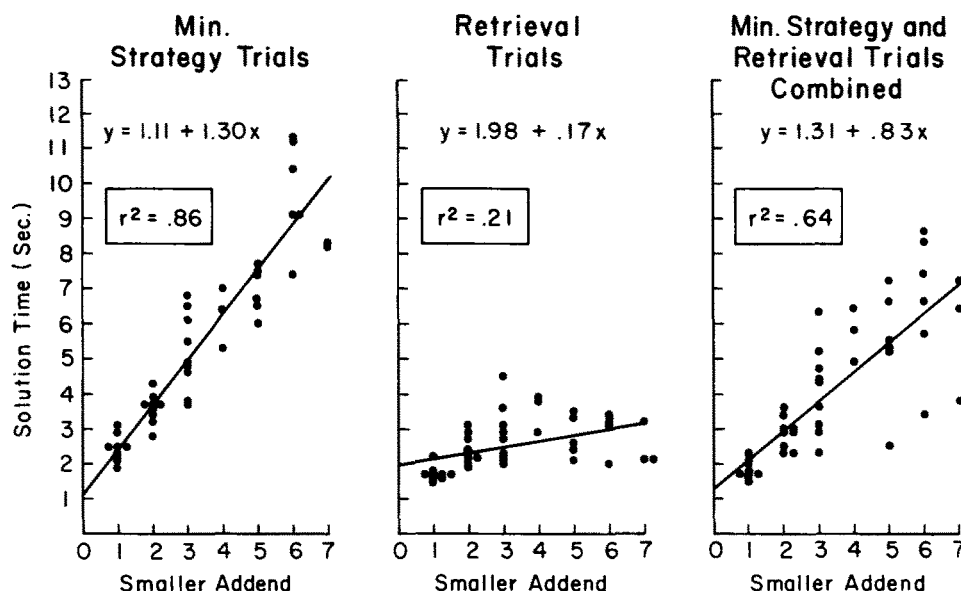


Figure 2. Relation between size of the smaller addend and median solution time on each problem on trials in which children reported using the min strategy (left panel), retrieval (center panel), or either strategy (right panel).

Table 5
Best Predictor of Unconditional and Conditional Probabilities of Use of Each Strategy on Each Problem

Strategy	Unconditional probabilities		Conditional probabilities	
	Best predictor	R^2	Best predictor	R^2
Retrieval	Sum	70	Sum	70
Decomposition	Bigger10	54	Bigger10	58
Min	Sum ²	57	Small min/big diff	79
Count all	Difference	59	Sum ²	74
Guessing	Sum ²	65	Inapplicable	—

Note. If CP = conditional probability, $CP(\text{retrieval}) = P(\text{retrieval})$; $CP(\text{decomposition}) = P(\text{decomposition})/[1 - P(\text{retrieval})]$; $CP(\text{min}) = P(\text{min})/[1 - P(\text{retrieval}) - P(\text{decomposition})]$; and $CP(\text{count all}) = P(\text{count all})/[1 - P(\text{retrieval}) - P(\text{decomposition}) - P(\text{min})]$.

The reasons why size of the smaller addend so accurately predicted percentage of errors aggregated across strategies were the same as those underlying its success in predicting the aggregated solution times. Retrieval and the min strategy were by far the most often used approaches. The min strategy produced greater variability than retrieval in percentage of errors across problems ($SD = 13$ vs. 10 for percentage of errors on retrieval trials). The size of the smaller addend accounted for 74% of variance in percentage of errors on min strategy trials, whereas no variable accounted for more than 30% of variance in percentage of errors on retrieval trials. Finally, the size of the smaller addend correlated positively with percentage of errors when approaches other than the min strategy were used. Its correlations with percentage of errors on retrieval, decomposition, counting-all, and guessing trials were .55, .33, .46, and .27, respectively.

Strategy Use

For each strategy, two regression analyses were conducted to determine the types of problems on which children most often used that strategy. In one, the dependent variable was the absolute (unconditional) probability of using the strategy on each problem. In the other, the dependent variable was the conditional probability of using the strategy on each problem, given that a faster strategy was not used on it. The formulas for the conditional probabilities are presented at the bottom of Table 5. It should be noted that conditional probabilities of guessing were treated somewhat differently than were those of the other approaches. Guessing was viewed as not being a strategy in the same sense as the other approaches, but rather as representing a default option, to be invoked here and elsewhere when no other approach seems likely to yield a correct answer. Therefore, conditional probability of guessing was viewed as being dependent on all other strategies not being used.

The upper portion of Table 5 shows the results of the analyses. For simplicity, only the best predictor of use of each strategy is presented. As discussed in the introduction, each strategy was expected to be used most often on problems in which it was easiest to execute, either in absolute terms or relative to other

strategies. This view suggested that the best predictor of use of retrieval would be the size of the sum or the squared sum. In fact, the best predictor of its use was the sum, which accounted for 70% of the variance in its frequency of use on the 45 problems. The size of the squared sum was an almost equally good predictor, accounting for 68% of the variance. In both cases, the smaller the sum (sum squared), the more often retrieval was used. Given the extremely strong correlation between the sizes of the sum and the squared sum ($r = .97$), it is unclear that differences between these two predictors are meaningful. (Because retrieval was the fastest strategy, its conditional and unconditional probabilities of use were identical.)

The nature of the decomposition strategy suggested that it would be easiest to execute when one of the addends exceeded 10. Whether one of the addends exceeded 10 proved to be the best predictor of use of this strategy within the analyses of both unconditional and conditional probabilities. The variable accounted for slightly more variance in the analysis of the conditional probabilities than of the unconditional ones, 58% vs. 54%.

The best predictor of use of the min strategy was expected to be the size of the minimum addend, the size of the difference between the addends, or both. Here, the results of the analyses of conditional and unconditional probabilities diverged. The analysis of the conditional probabilities yielded results in accord with the prediction. Given that they did not use a faster strategy on the problem, children used the min strategy most often on problems that included a small minimum addend, a large difference between the addends, or both. This variable accounted for 79% of the variance in the data.

The results of the analysis of unconditional probabilities of use of the min strategy required a less straightforward interpretation. The *greater* the sum squared, the more often the min strategy was used. This variable accounted for 57% of the variance. The only obvious interpretation of this result is that retrieval was used very often on problems with small squared sums, thereby limiting the potential use of the min strategy on them. (Note that this interpretation, emphasizing the impact of frequency of the faster retrieval strategy on the frequency of the slower min strategy, is essentially identical to the argument for conditionalizing probability of use of each strategy on not using faster strategies on the problem.)

The best predictor of use of the counting-all strategy was expected to be the amount of counting that needed to be done, that is, the sum. Again, the analyses of the conditional and unconditional probabilities diverged. The analysis of conditional probabilities yielded results in accord with the prediction. Given that a faster strategy was not used on a problem, the sizes of the sum squared and sum were the best predictors of how often counting all was used, accounting for 74% and 72% of the variance, respectively. The smaller the sum and sum squared, the more likely children were to use the counting-all strategy.

As with the min strategy, the analysis of unconditional probability of use of the counting-all strategy required a less straightforward interpretation. The best predictor was the difference between addends, which accounted for 59% of the variance. The smaller the difference between addends, the more often the counting-all strategy was used. This initially surprising pattern

may be due to the faster min strategy being used most frequently on problems with large differences between addends, thereby limiting use of counting all to problems with small differences between addends. The pattern again suggests that faster strategies have an asymmetric influence on when slower strategies are used.

Children were hypothesized to guess or not advance any response when they could not execute any of the other strategies accurately. The results were consistent with this view. Children most often guessed or said that they did not know the answer on problems with large squared sums and sums. The two variables accounted for 65% and 62% of the variance, respectively. No parallel analysis of unconditional probabilities could be conducted because the conditional probability of guessing was equal to 1.00 on all problems.

Discussion

The basic findings of this study can be summarized quite simply. When performance was averaged over trials, results replicated those reported previously. When performance produced by each strategy was analyzed separately, however, a different picture emerged. Rather than the min strategy or any other approach predominating, children seemed to use a variety of strategies in ways that produced adaptive combinations of accuracy and speed. Overall, children's addition procedures seem to be considerably more varied, flexible, and finely attuned to task demands than portrayed in previous models of children's arithmetic.

These data place a number of constraints on future models of the development of addition skills. Such models must portray children as using a number of strategies throughout the early elementary school years (as well as during the preschool period, cf. Siegler & Shrager, 1984). This must be true within as well as across individuals. It also must be true within as well as across problems; children cannot be depicted as always using retrieval on some problems, such as ties, and always using counting strategies on other problems. The distribution of strategies must gradually shift from more use of counting all and guessing to more use of retrieval and decomposition. The min strategy must be used at all ages, but must not be used on all or almost all trials at any age. Each strategy's speed and accuracy of execution must improve with age and experience. Finally, the frequency with which each strategy is used must reflect both its and other strategies' speed and accuracy on particular problems.

The findings have implications not only for understanding children's arithmetic but also for understanding a number of more general conceptual and methodological issues relating to people's use of multiple strategies. Among these are the relation between averaging over strategies and averaging over individuals, the types of tasks on which people are most likely to use multiple strategies, the role of verbal protocols in assessing strategy use, and the processes by which strategies are chosen.

Averaging Over Individuals Versus Averaging Over Strategies

In previous analyses, the problems raised by averaging over strategies have been equated with those raised by averaging over

subjects. At times, the problems raised by the two types of averaging are identical: when each subject in an experiment consistently uses a single strategy, but different subjects use different strategies. This is the case envisioned by Newell (1973) in his analysis of the dangers of averaging over strategies.

If individuals use multiple strategies, however, the two potential sources of distortion become separate problems. Indeed, under these conditions, analyses of individual performance can encounter the same difficulties as analyses of group-level performance. This has clearly been the case in children's addition. As was noted earlier, chronometric analyses of individual performance have yielded results similar to chronometric analyses of group performance. In both, the min model fit the data quite well. Yet it appears that few children consistently use the min strategy. Although analyses of individual subjects' performance are indispensable, analyses of performance generated by individual strategies seem equally essential.

When Do People Use Multiple Strategies?

The phenomenon of individuals using several strategies extends far beyond simple arithmetic. Some of the tasks on which it already has been documented are spelling, reading, decision making, telling time, referential communication, use of past-tense verbs, solving missionaries-and-cannibals problems, series completion, and number conservation (e.g., Ilg & Ames, 1951; Jorm & Share, 1983; Kahan & Richards, 1986; Lefevre & Bisanz, 1986; Maratsos, 1983; Payne, 1976; Shultz, Fisher, Pratt, & Rulf, 1986; Siegler, 1981; Siegler & McGilly, in press; Simon & Reed, 1976). One factor that seems to influence whether people use multiple strategies is their degree of knowledge about the task. Multiple-strategy use has been most evident at points in the acquisition process where people have had sufficient experience with a task to have learned several ways of solving problems, but not so much that they always use a single, efficient method. Even in the restricted age range examined in the present study, the older children had stopped using the counting-all approach that was used frequently by the younger children. Presumably, by later childhood the min strategy and decomposition also would cease being used on these simple problems.

More speculatively, tasks on which situational and individual differences in strategy use have been documented may often be ones in which individuals are also using multiple strategies. Many of the differences between individuals and situations may prove to be ones in the extent of use of particular strategies, rather than all-or-none differences. For example, people high in spatial ability have been depicted as using different strategies from those used by people lower in spatial ability on spatial comparison, mental rotation, sentence verification, and transitive inference tasks (Cooper & Mumaw, 1985; Just & Carpenter, 1985; Mathews, Hunt, & McLeod, 1980; Sternberg & Weil, 1980). Further analyses may reveal that both types of individuals use both types of strategies, but with differing frequency and perhaps under different ranges of conditions.

How Can Multiple Strategy Use Be Assessed?

In many situations, simply asking people immediately after each response how they generated their answers may be the best

way to separate data produced by different strategies. Ericsson and Simon (1984) noted that when people are asked to report on immediately previous trials, they often can accurately retrieve considerable episodic information, particularly when the processing episode was not extremely brief in duration.

The very success of indirect methods of cognitive assessment, such as chronometric analyses, error analyses, and eye-movement analyses, may have led to an underuse of verbal reports as sources of data. As Nisbett and Wilson (1977) noted, verbal reports can convey seriously misleading pictures of what people are doing. However, as the present results demonstrate, analyses of solution time and error data, even ones that seem very successful, can communicate equally misleading pictures. On tasks in which people can give coherent descriptions of how they generated the immediately previous response, it seems well worthwhile to obtain the self reports and to see if they converge with, and prove revealing about, solution times, errors, and other indirect indicators of strategy use.

The situation is more difficult when subjects cannot give valid verbal reports. Even here, however, variability in strategy use need not simply be accepted as a source of error variance (Farah & Kosslyn, 1982). Probably the most basic approach is to perform a task analysis, imagine possible strategies for performing the task, and examine the raw data to see if subsets fit different strategies. For example, in the present study, solution times tended to fall into two clusters at high values of the smaller addend. Most of the long times were generated by the min strategy, most of the short ones by retrieval. Examining particular errors can also prove revealing. On tasks in which multiple errors are possible, the failure of the hypothesized strategy to easily generate common errors provides grounds for suspecting that other strategies also are being used. The identity of the other strategies may be inferred by considering what approaches would generate the particular incorrect answers. Priming use of different strategies by first presenting problems thought likely to elicit use of one strategy or another and then contrasting error and solution time patterns on target problems may also prove useful. Finally, on some tasks people may provide revealing verbal reports if probed during the solution process, although they would be unable to do so by the end of the trial (Staszewski, 1987). None of these methods is as easy to apply and interpret as full verbal reports provided immediately after task performance. However, each of them may aid detection of multiple strategies in some situations where accurate verbal reports cannot be obtained.

How People Choose Strategies

Several general lessons of the study concern when and how people choose particular strategies. One is that the conditions under which each strategy is adopted must be viewed in the context of the total set of available strategies. Even if a strategy produces its fastest and most accurate performance on particular problems, it will not be used often on them if other strategies produce equally accurate and faster performance. For example, in the present study the min strategy produced its fastest and most accurate performance on problems in which both addends were small, such as $4 + 1$ and $5 + 2$. In absolute terms,

however, the strategy was used on only 24% of trials on the 9 small addend problems, which is well below its average of 36% across all 45 problems. Retrieval produced equally accurate, and considerably faster, performance on these problems and therefore was usually used on them.

The interactive nature of strategy choices also was evident in the best predictors of the absolute probability of use of each strategy. The min strategy was used most often on problems that had large squared sums, presumably because the faster retrieval strategy could be used accurately on problems with small squared sums. The counting-all strategy was used most often on problems with small differences between addends, presumably because the faster min strategy was used most often on problems with large differences between addends. Guessing was used most often on problems with large sums, presumably because two faster strategies, retrieval and counting all, were used most often on problems with small sums. Thus, when a strategy is used will reflect not only the conditions under which it is most effective but also the conditions under which other strategies are.

Both speed and accuracy seem to influence each strategy's conditions of use, but they do so in different ways. Fast strategies are chosen when they can generate accurate performance. Slower strategies are chosen when they can generate accurate performance and faster ones cannot. This pattern was observed previously in studies focusing on how people choose between stating a retrieved answer and using a backup strategy (Siegler, 1986; Siegler & Shrager, 1984). The present study replicated that finding and revealed a parallel pattern in the more general case of choices among multiple nonretrieval strategies as well as retrieval. It also indicated that conditionalizing probability of strategy use on not using a faster strategy can reveal the precise ways in which the accuracy of a strategy influences when the strategy is used. Given that faster strategies were not used, counting all and retrieval were used primarily on problems with small sums and squared sums, decomposition was used primarily on problems with an addend that exceeded 10, and the min strategy was used primarily on problems that had either small minimum addends or large differences between addends. The findings with the min strategy in particular suggest that both the absolute accuracy of a strategy and its accuracy relative to other strategies influence choices among strategies. The conditional probability of using the min strategy was high not only where it would be easy to use in absolute terms, problems with small minimum addends, but also where its advantage relative to the counting-all and retrieval strategies would be greatest, problems with large differences between addends.

Development of Strategy Choices

The present findings indicated that the speed and accuracy with which children execute each addition strategy increase with age and experience. Accompanying this change is greater use of the two fastest strategies, retrieval and decomposition, and decreasing use of the slowest strategy, counting all. Beyond these general trends in strategy use, however, lie a large number of unanswered questions about how strategy choices develop.

One concerns the mechanisms responsible for use or nonuse

of a strategy. Despite the greater efficiency of the min strategy relative to the counting-all approach (it was as fast or faster and as accurate or more accurate on literally every problem where both strategies were used), 55% of kindergarteners reported using both procedures. In contrast, only 11% of first graders reported using counting all as well as the min strategy, and no second grader did. What processes lead to temporary continuation of use of globally less efficient strategies when more efficient strategies are also known and used? What processes lead to the eventual abandonment of the less efficient strategies?

A related issue concerns the level at which information about each strategy is represented. As people gain experience with a new strategy, do they represent information about its speed and accuracy as a general characteristic of the strategy, as a relation between the strategy and particular problems, as a relation between the strategy and classes of problems, or at several of these levels? For example, if children just acquiring the min strategy learned that it yielded faster and more accurate performance than the counting-all strategy on $13 + 2$ and $13 + 3$, would they then use the min strategy more often only on those two problems, more often on all addition problems, or more often the more similar the target problem to the original problem (according to some generalization gradient)?

A further issue concerns acquisition of new strategies. How do people learn new strategies such as decomposition? How do they integrate them with their existing organization of strategies? How do they modify them to fit new situations? Taking the diversity of strategy use seriously not only challenges conclusions based on averaged data, it raises whole new sets of issues.

References

- Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, 2, 213-236.
- Ashcraft, M. H. (in press). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology*. New York: Springer-Verlag.
- Ashcraft, M. H., & Battaglia, J. (1978). Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 527-538.
- Ashcraft, M. H., Fierman, B. A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review*, 4, 157-170.
- Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75-112). Hillsdale, NJ: Erlbaum.
- Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9-24). Hillsdale, NJ: Erlbaum.
- Cooper, L. A., & Mumaw, R. J. (1985). Spatial aptitude. In R. Dillon (Ed.), *Individual differences in cognition* (pp. 67-94). Orlando, FL: Academic Press.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge: MIT Press.
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134-140.
- Farah, M. J., & Kosslyn, S. M. (1982). Concept development. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 125-170). New York: Academic Press.
- Fuson, K. C. (1982). An analysis of the counting-on solution procedure in addition. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 67-81). Hillsdale, NJ: Erlbaum.
- Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, 79, 329-343.
- Groen, G., & Resnick, L. B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology*, 69, 645-652.
- Hamann, M. S., & Ashcraft, M. H. (1985). Simple and complex mental addition across development. *Journal of Experimental Child Psychology*, 40, 49-72.
- Houlihan, D. M., & Ginsburg, H. P. (1981). The addition methods of first- and second-grade children. *Journal for Research in Mathematics Education*, 12, 95-106.
- Ilg, F., & Ames, L. B. (1951). Developmental trends in arithmetic. *Journal of Genetic Psychology*, 79, 3-28.
- Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics*, 4, 103-147.
- Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, 92, 137-173.
- Kahan, L. D., & Richards, D. D. (1986). The effects of context on children's referential communication strategies. *Child Development*, 57, 1130-1141.
- Kaye, D. B. (1986). The development of mathematical cognition. *Cognitive Development*, 1, 157-170.
- Kaye, D. B., Post, T. A., Hall, V. C., & Dineen, J. T. (1986). The emergence of information retrieval strategies in numerical cognition: A developmental study. *Cognition and Instruction*, 3, 137-166.
- LeFevre, J., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition*, 14, 287-298.
- Maratsos, M. P. (1983). Some current issues in the study of the acquisition of grammar. In P. H. Mussen (Ed.), *Manual of child psychology* (Vol. 3, pp. 707-786). New York: Wiley.
- Mathews, N. N., Hunt, E. B., & MacLeod, C. M. (1980). Strategy choice and strategy training in sentence-picture verification. *Journal of Verbal Learning and Verbal Behavior*, 19, 531-548.
- Newell, A. (1973). You can't play 20-questions with nature and win: Projective comments on the papers of the Symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 463-526). New York: Academic Press.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90-138.
- Secada, W. G., Fuson, K. C., & Hall, J. W. (1983). The transition from counting-all to counting-on in addition. *Journal for Research in Mathematics Education*, 14, 47-57.
- Shultz, T. R., Fisher, G. W., Pratt, C. C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, 57, 143-152.
- Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 46, 1-74.
- Siegler, R. S. (1986). Unities across domains in children's strategy

- choices. In M. Perlmutter (Ed.), *Minnesota symposium on child development* (Vol. 19, pp. 1-48). Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & McGilly, K. (in press). Strategy choices in time telling. In I. Levin & D. Zakay (Eds.), *Psychological time: A life span perspective*.
- Siegler, R. S., & Shrager, J. (1984). A model of strategy choice. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.
- Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem-solving task. *Cognitive Psychology*, 8, 86-97.
- Staszewski, J. (1987). *Psychological reality of retrieval structures*. Unpublished doctoral dissertation, Cornell University, Ithaca, New York.
- Sternberg, R. J., & Weil, E. M. (1980). An aptitude $\times$ strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology*, 72, 226-239.
- Suppes, P., & Groen, G. (1967). Some counting models for first-grade performance data on simple addition facts. In J. M. Scandura (Ed.), *Research in mathematics education* (pp. 34-43). Washington, DC: National Council of Teachers of Mathematics.
- Svenson, O. (1975). Analysis of time required by children for simple additions. *Acta Psychologica*, 39, 289-302.
- Svenson, O., & Broquist, S. (1975). Strategies for solving simple addition problems: A comparison of normal and subnormal children. *Scandinavian Journal of Psychology*, 16, 143-151.
- Svenson, O., Hedenborg, M.-L., & Lingman, L. (1976). On children's heuristics for solving simple additions. *Scandinavian Journal of Psychology*, 20, 161-173.

Appendix

Factors That Imperil Analyses of Data Averaged Over Strategies

The influence on data averaged over strategies of relative frequency of strategy use, variability of performance generated by each strategy, and independent-dependent variable relations across and within strategies can be illustrated unambiguously through analyses of synthetic data. In all of the analyses, the strategies being compared are the min strategy and retrieval, and the problem set is the one studied by Groen and Parkman (1972), the 55 problems with sums between 0 and 9 inclusive. Although the illustrations specifically involve children's addition, the three sources of variation would influence analyses on any task where people used two or more strategies.

The first synthetic analysis is a base case, intended to provide a point of comparison for the explicit manipulations of variables in the other analyses. As is shown in Table A-1, in this base case, frequency of use of the two strategies on each problem is identical; variances generated by the two strategies are similar; solution times generated by the two strategies are moderately correlated; and the predictor variable associated with each strategy (sum squared with retrieval; smaller addend size with the min strategy) is perfectly correlated with the solution times when that strategy is used. Under these base-case conditions, the empirical predictors associated with the two strategies accounted for similar percentages of variance in the averaged data, 72% and 76%. Note that despite the fact that each strategy was *not* used on 50% of trials on each problem, the predictors associated with both of them provided excellent accounts of the averaged solution times, accounting for a sizable majority of variance in each case. The positive correlation between the times generated by each strategy, $r = .48$, meant that the predictor associated with each strategy accounted for times generated by the other strategy moderately well, as well as accounting for its own times perfectly.

The effects of frequency of use of each strategy can be seen by contrasting this base case with cases where one strategy was used on two thirds of trials and the other on one third. To simulate the case in which the min strategy was used on two thirds of trials and retrieval on one third, the solution time for each problem was an average of the times generated on three presentations of the problem. Two times were those projected for use of the min strategy and one was the time projected for use of retrieval. As is shown in Table A-1, this led to the predictor associated with the min strategy, the size of the smaller addend, becoming a substantially better predictor than the predictor associated with retrieval, the sum squared. Changing the procedure so that on each problem, two of the times were those projected for use of retrieval and one was that projected for use of the min strategy had the opposite effect.

Similar analyses can be used to illustrate the effects of different strategies producing different variability in dependent variable scores. Relative variability produced by the min strategy and retrieval was manipulated by varying the time-per-count of the min strategy. Longer times-per-count when using the min strategy would result in this strategy producing more variable solution times on different problems. As is shown in Table A-1, when the time-per-count for the min strategy was increased from the 1 s/count base rate to 2 s/count, the predictor associated with the min strategy, the size of the smaller addend, was the best

Table A-1
Analyses of Synthetic Data

Analysis	R_{Min}^2	R_{Sum}^2	SD_{min}	SD_{sum}
Base case	72	76	1.22	1.34
Frequency manipulated				
67% min, 33% retrieval	86	58	1.22	1.34
67% retrieval, 33% min	53	90	1.22	1.34
Variance manipulated				
Increments = 2 s/min, .05 s/sum ²	86	58	2.44	1.34
Increments = .5 s/min, .05 s/sum ²	53	90	0.61	1.34
Extreme case				
33% min, 67% retrieval	97	40	2.44	.27
Increment = 2 s/min, .01 s/sum ²				
Independent-dependent relations manipulated				
Retrieval times randomly assigned ^a	49	23	1.22	1.34
Min times randomly assigned ^b	13	53	1.22	1.34

Note. Unless otherwise stated, frequency on each problem = 50% min strategy and 50% retrieval; times on min trials = 1 s/min; times on retrieval trials = .05 s/sum²; average time on each problem = (time_{min} + time_{retrieval})/2; correlation of times generated by min strategy and retrieval, $r = .48$; correlation between predictor variable and dependent variable within each strategy, $r = 1.00$.

^a r retrieval times-min times = .09; r retrieval time-sum² = .03. ^b r min times-retrieval times = .07; r min times-smaller addend = .06.

predictor of the averaged solution times. When the time-per-count of the min strategy was reduced to .5 s/count the predictor associated with retrieval, the sum squared, was the best predictor of the averaged times.

If the variability in solution times produced by two strategies is sufficiently discrepant, the predictor associated with the less frequently used strategy can account for *more* variance when the data are averaged across strategies. The extreme case in Table A-1 illustrates this point: Retrieval is used twice as often as the min strategy, the increment per unit of sum squared is .01 s, and the increment per unit of the smaller addend is 2 s. Despite the fact that the min strategy is used on only one third of trials, the size of the smaller addend accounts for the pattern of averaged solution times very well; despite the fact that retrieval is used on two thirds of trials, the size of the sum squared does not account for the pattern nearly as well. Large discrepancies in the variability of performance generated by different strategies are not just a hypothetical possibility. In the present experiment, the standard deviation of children's median solution time on each problem was almost four times as great for min strategy trials as for retrieval trials (2.72 vs. .73).

The role of independent-dependent variable relations across and within strategies can be illustrated in a similar way. The demonstration began with the same 55 times generated by each strategy as in the base case. The manipulation involved randomly assigning to problems the times generated by one of the strategies before computing the mean time (averaged across the two strategies) for each problem. This left the

frequency of use and the variance generated by each strategy the same as in the base case. However, the manipulation had large effects on the relation between predictor variables and the averaged data. Not surprisingly, the manipulation greatly reduced the correlation with the averaged times of the predictor associated with the strategy whose times were randomly reassigned. The predictor-dependent variable relation for the times generated by this strategy had been eliminated. More striking was the degree to which the random reassignment of times produced by one strategy decreased the correlation between the predictor associated with the other strategy and the averaged times. For example, randomly reassigning to problems the times generated by *retrieval* reduced the percentage of variance that could be explained by the predictor associated with the *min strategy* (the size of the smaller addend) from 72% to 49%.

In sum, relative frequency of use of each strategy, relative variance generated by each strategy, and relations of independent and dependent variable values within and across strategies all influence analyses of data averaged over strategies. Together, they can lead to a model giving an excellent account of data averaged over strategies, even if the strategy presumed by the model is not used very often.

Received November 4, 1986

Revision received February 26, 1987

Accepted March 5, 1987 ■