Hazards of Mental Chronometry: An Example From Children’s Subtraction

Robert S. Siegler
Carnegie Mellon University

Chronometric analyses have frequently been used to infer the strategy people use to solve classes of problems. However, in the many situations in which people use multiple strategies, chronometric analyses can lead to incorrect conclusions about strategy use. This was illustrated in the context of 2nd and 4th graders’ simple subtraction. As in previous studies, Woods, Resnick, and Groen’s (1975) chronometric model was found to fit the pattern of solution times averaged over strategies. Strategy classifications based on observation of overt behavior and immediately retrospective verbal reports, however, indicated that (a) children rarely used 1 of the 2 approaches posited by the model, (b) they used the other approach no more often on problems where they were posited always to use it than on problems where they were posited never to use it, and (c) they used a greater variety of strategies than posited by the Woods et al. model. The findings raised issues about when chronometric methods can be usefully applied, how strategy use can be assessed, and how people choose among multiple strategies.

Chronometric analyses, in which solution time patterns are used to infer mental processes, have been frequently and profitably applied to studying reading, mathematics, and other topics of educational interest (e.g., Just & Carpenter, 1987; Mayer, 1987; Sternberg, 1985). Used as the sole index of cognitive activity, however, such analyses can yield seriously distorted views of the processes they are intended to identify. The problem is particularly serious when people use diverse strategies to solve a particular type of problem, rather than always using a single one. Under such circumstances, trying to infer the strategy that is being used can only lead to incorrect conclusions. This point was illustrated in Siegler (1987a), where an addition strategy that appeared to account for a large body of chronometric data was found actually to be used on only about one third of trials. The present study reports an even more extreme case. A model of children’s subtraction that fit a considerable body of chronometric data was found never to be used at all.

The purpose of this article is not just to demonstrate that a previously proposed model of children’s subtraction is wrong. It is also to argue for a fundamentally different view of the basic phenomena in this and many other academic domains. Within this view, the basic phenomena are (a) that people often use a wide variety of strategies, even in solving a single problem; (b) that this diversity of strategy use holds true within, as well as between, subjects; (c) that changes with age and experience occur in the distribution of strategies as well as in the speed and accuracy associated with each one; and (d) that people often choose among strategies in ways that result in each strategy’s being used most often on problems where the strategy’s speed and accuracy are advantageous, relative to those of other available procedures. Methods for studying academic skills must reflect, not obscure, these phenomena. The present experiment illustrates both how conventional chronometric analyses can obscure what people are doing and how alternative analyses can reflect thought processes more faithfully.

Previous Research on Children’s Subtraction

Existing research suggests two strikingly different views of elementary-school children’s subtraction. One view is based primarily on chronometric data, the other on self-report data. First, consider the type of model that has emerged from chronometric analyses. The most prominent chronometric model of elementary-school children’s subtraction is Woods, Resnick, and Groen’s (1975) smaller-count model. Within this model, children always execute whichever of two counting strategies can be executed with fewer counts on the particular problem. One type of count involves counting down from the minuend (first number) the number of counts indicated by the subtrahend (second number) and stating as the answer the number at which counting stops. Thus, on 12 – 3, children would count down from 12 to 9 and advance 9, the stopping point, as the answer. The other approach involves counting up from the second number to the first one and stating the answer. The number of counts required to reach that point is the number of counts to be subtracted. For example, on 12 – 9, children would count up from 9 to 12 and advance 3, the number of counts, as the answer. A child who is behaving in accord with the smaller-count model would always use the counting down strategy on 12 – 3, because it can be executed with 3, rather than 9, counts. The child would always use the counting up approach on 12 – 9 for the same reason.

This research was supported in part by National Institutes of Health Grant HD-19011, in part by a grant from the Spencer Foundation, and in part by a grant from the McDonnell Foundation. 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.

Correspondence concerning this article should be addressed to Robert S. Siegler, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890.
The smaller-count model is consistent with a substantial body of chronometric data. Woods et al. (1975) reported that the model accounted for roughly 50% of the variance in both second and fourth graders' solution times on simple subtraction problems. Svenson and Hedenborg (1979) found that it accounted for a similar percentage of the variance in Swedish third graders' solution times on a somewhat more difficult problem set. Both Woods et al. and Svenson and Hedenborg found that it predicted more effectively than other models for the large majority of individual children, as well as for the group as a whole. For example, in the Woods et al. study, it was the best predictor of solution times for 30 of 40 second graders and for all 20 fourth graders. Groen and Poll (1973) found that the smaller-count model also explained roughly 70% of the variance in solution times on open sentence addition problems such as 4 + ? = 7, which, like subtraction problems, can be solved either by counting up or counting down.

In contrast to this smaller-count model, in which children are depicted as always counting up on some problems and always counting down on others, evidence from verbal reports suggests that children use a wide range of strategies both within and across problems. Even within a single problem, young elementary-school children report not only counting down and counting up but also retrieving answers, referring to related addition and subtraction problems, and decomposing a single problem into two simpler problems (Carpenter, 1985; Carpenter & Moser, 1984; Fuson, 1984; Ilg & Ames, 1951; Siegler, 1987b; Steinberg, 1983).

Reconciling Findings From the Two Approaches

How can the good fit of the smaller-count model to the chronometric data be reconciled with the self-reports of diverse subtraction strategies? One possibility is that the verbal self-reports are epiphenomenal. People in general often cannot accurately report their cognitive processes (Nisbett & Wilson, 1977); the problem may be even greater with children (Brainerd, 1973). In the specific case of subtraction, children may count quickly but be unaware of what they are doing and therefore simply guess at their method. Previous studies that have reported protocol data have not presented converging evidence to validate the verbal explanations. Thus, children's self-reports may not correspond to how they actually solved the problems.

The other possibility is that children's explanations do accurately reflect their subtraction strategies and that the fit of the smaller-count model to the chronometric data is misleading. A recent study of kindergarteners', first graders', and second graders' addition (Siegler, 1987a) presented a potentially analogous case that illustrates how this can happen. Just as the smaller-count model is consistent with a considerable amount of chronometric data in subtraction, Groen and Parkman's (1972) min model was consistent with a large amount of data in addition. Within the min model, children always solve simple addition problems by counting up from the larger addend the number of counts indicated by the size of the smaller addend. For example, a child using the min strategy would solve 3 + 6 by thinking "6, 7, 8, 9". Within this model, the only factor that influences solution times is the amount of counting up from the larger addend. The time per count is constant. Therefore, the model predicts that solution times will increase linearly with the size of the smaller addend.

Consistent with this prediction, Groen and Parkman (1972) found that the size of the smaller addend was the best predictor of solution times on simple addition problems. It was a very good predictor in absolute terms as well, accounting for roughly 70% of the variance in solution times. Numerous other investigators also found that the size of the smaller addend was the best predictor of solution times and an excellent predictor in absolute terms (Ashcraft, 1982; Geary, Widaman, Little, & Cormier, 1987; Goldman, Pellegrino, & Mertz, 1988; Groen & Resnick, 1977; Kaye, Post, Hall, & Dineen, 1986; Svenson, 1975). These studies included diverse problem sets, European as well as North American children, learning disabled as well as typical children, and analyses of individual children as well as groups (Geary et al., 1987; Goldman et al., 1988; Groen & Resnick, 1977; Kaye et al., 1986; Svenson & Broquist, 1975).

Children's verbal self-reports suggested a quite different picture than the solution time data. When asked how they had solved addition problems, children reported using not only the min strategy but counting from one, retrieving answers from memory, and decomposing problems into two simpler problems, among other approaches (Carpenter & Moser, 1982; Fuson, 1982; Houlihan & Ginsburg, 1981; Ilg & Ames, 1951).

To test whether such verbal reports were misleading or whether the chronometrically based models were, Siegler (1987a) obtained both solution times and verbal reports from each child on each trial. When data were averaged over all trials (and over all strategies) as in earlier studies, the results closely replicated the previous finding that solution times were a linear function of the smaller addend. Size of the smaller addend accounted for 76% of the variance in solution times on different problems. If these analyses had been the only ones conducted, the usual conclusion would have been reached, namely that first and second graders consistently use the min strategy to add.

The children's verbal reports, however, suggested a quite different picture. The min strategy was but one of five approaches that children reported using. On only 36% of trials did they say that they had counted from the larger addend. This reporting of diverse strategies characterized individual as well as group performance; most children reported using at least three strategies.

Dividing the solution time data according to what strategy children said they had used on that trial lent considerable credence to the children's verbal reports. 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 Siegler (1987a) data set as a whole; it accounted for 86% of the variance in solution times. In contrast, on trials where 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. It never accounted for as much as 40% of the variance and was never either the best or
the second-best predictor of solution times. A variety of data converged on the conclusion that children used the five strategies that they reported using and that they used them on those trials where they said they had.

Implications

These findings, together with the parallels between the addition and subtraction literatures, suggested that it would be useful to employ a similar research strategy to study children’s subtraction. Both solution times and immediately retrospective self-reports of strategy use were collected on each trial. Then the solution times were analyzed both for the entire data set and separately, according to the strategy children said they had used on the trial. If children do in fact consistently use the smaller-count strategy and their verbal reports are epiphenomenal, the predictors associated with the smaller-count strategy should predict not only solution times averaged over all trials but also solution times separated according to the strategy that children reported using. In contrast, if children’s self-reports are accurate and the averaged solution times are misleading, then separate analyses of solution times generated by each strategy should yield distinct patterns that correspond to a priori sources of difficulty of executing the particular strategy. For example, the size of the smaller number would be expected to be a very good predictor of solution times on trials where children reported counting down the number of times indicated by the smaller number. However, it would not be expected to be a particularly good predictor on trials where children reported decomposing the problem into two simpler ones or referring to a related addition problem. If this latter result emerges, a further task will be to explain why the chronometric model fits the averaged data if children do not use the hypothesized strategy.

Method

Participants

The children were 37 second graders and 37 fourth graders at a middle-class suburban school. The mean chronological age of the second graders was 90 months; that of the fourth graders was 117 months. Both second-grade and fourth-grade samples included 20 girls and 17 boys. Children in both grades had received substantial experience in their classrooms with the types of subtraction problems that were presented. A 34-year-old female research assistant was the experimenter.

Problems

Children were presented with 36 problems, 9 each of four types. Each type of problem included the nine factorially possible combinations of subtrahend (smaller number) size and difference size. One type of problem involved large minuends (13–17), small subtrahends (1–3), and large differences (12–14). A second type of problem involved large minuends (13–17), large subtrahends (12–14), and small differences (1–3). A third type of problem involved large minuends (13–17), medium subtrahends (8–10), and medium differences (5–7). Finally, the fourth type of problem involved small minuends (5–9), small subtrahends (1–3), and medium differences (4–6).

These problems were chosen so as to include items in which counting up was much easier than counting down (e.g., 15 – 13), in which counting down was much easier than counting up (e.g., 15 – 2), and in which the number of counts was approximately equal (e.g., 15 – 8). The problems were also chosen so as to provide some very familiar problems and other fairly unfamiliar problems, problems of a wide range of difficulty, and problems for which special strategies could be used (e.g., problems with 10 as the number to be subtracted).

Procedure

Each child was brought individually from the classroom to an empty room within the school. The child sat directly across the table from the experimenter. At the outset of the first session, the child was told:

We are going to do some subtraction 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 do the best that you can.

These instructions stressed the importance of accuracy rather than speed, because of a desire not to constrain the range of strategies that children used.

After the first problem the child was told, “We’re interested in how children your age figure out the answers to these problems. How did you figure out the answer to that problem?” This question was repeated after each item until children began volunteering the information; thereafter, the question was asked only when they did not spontaneously describe what they had done. If the description was unclear, the experimenter asked follow-up questions. For example, if the child said only, “I counted,” the experimenter asked, “What number did you start counting from? How did you count then?”

The 36 problems were presented to children over 2 school days. A randomly chosen set of 18 problems (different for each child) was presented each day. The days were as close to contiguous as possible; they almost always were consecutive school days. At the beginning of the second session, the experimenter reminded the child of the game they had played earlier; if children indicated that they remembered the previous session (which virtually all did), the experimenter said simply that they would play the same game again.

Each child’s behavior was recorded with a Sony SLO-325 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 0.1 s which seemed to be a sufficient degree of accuracy for the present task, in which the median solution time was 4.1 s.

When children produced overt behavior, the videotaped record of that behavior was the main guide to strategy classification. When they did not, the verbal protocols were used to classify performance. Together, the explanations and the overt problem-solving activities allowed highly reliable classification of each child’s strategy on each trial. Two independent raters agreed on 92% of their classifications. In those 8% of cases where they initially disagreed, they discussed the observations and reviewed the tape until they reached agreement. Agreement on the initial classifications of each strategy was at least 85%.

Results

Overview

A few summary statistics may provide an overall sense of the children’s level of performance. Children answered co-
rectly on 84% of trials; second graders were correct on 74%, and fourth graders were correct on 94%. Median solution time on each trial was 4.1 s, 7.2 s for second graders and 2.3 s for fourth graders. As shown in Table 1, children appeared to use six main strategies: retrieval, counting down, counting up, deleting 10s, addition reference, and guessing or not responding. Retrieval, counting down, and counting up were described in the beginning of this article. Deleting 10s involved a type of decomposition in which children treated the 10s value separately from the 1s value. For example, on 15 - 3, they might explain their answer by saying, "5 - 3 = 2, and you put back the 1, so 12." Addition reference involved explaining an answer to a subtraction problem in terms of a related addition problem. For example, children sometimes explained their having said that 13 - 5 = 8 by saying "I knew that 5 + 8 = 13." The Table 1 percentages sum to less than 100% because only strategies that children used on at least 3% of trials are listed. Among the strategies that did not meet this criterion, the two most frequently used were imagining doing the problem on paper (2% of trials) and referring to related subtraction problems (1%).

The diversity of strategy use that is apparent in Table 1 was not simply a case of one child using one strategy and another child using a different one. Fully 99% of individual children used multiple strategies. The children's verbal reports and overt behavior indicated that 8% of children used two strategies, 19% used three, 31% used four, and 41% used five or more. Thus, more than 70% of children used at least four different strategies to solve this set of simple subtraction problems. Use of diverse strategies was apparent both among the second graders, 62% of whom used four or more strategies, and among the fourth graders, 81% of whom did. The diversity of strategies also was evident within individual problems. On all but the smallest problems, large numbers of children used each of at least four different strategies.

Strategies differed considerably in frequency of use, solution times, and accuracy. Across all children, retrieval and counting down from the larger number were the most common strategies; each was used on approximately 35% of trials. Strategy use differed considerably between the two age groups. Counting down, guessing, and not responding were more common among younger children than among older ones; retrieval and addition reference were more common among older than among younger ones (Table 1).

Solution times associated with use of the strategies also varied considerably (Table 1). Across the two grades, retrieval was by far the fastest strategy, addition reference and deleting 10s were the next-fastest, counting down and up were the next-fastest, and guessing was the slowest. Children became considerably faster with age in executing each of these strategies.

Strategies also differed in accuracy. For both second and fourth graders, retrieval was the most accurate strategy (97% and 98% correct for the younger and older children, respectively) followed by deleting 10s (85% and 94% correct), addition reference (74% and 93% correct), counting down (73% and 89% correct), counting up (72% and 84% correct), and finally guessing or no response (14% and 50% correct). The accuracy data provided strong convergent validity for the distinction between retrieval and guessing. Children were correct on only 21% of trials on which they said that they had guessed, but were correct on 97% of trials where they said they had remembered the answer or had just known it.

Clearly, children knew when they knew and knew when they didn't know.

Predictors of Solution Times

Analyses of times averaged over strategies In previous studies that have emphasized chronometric data, multiple regression analyses of mean or median solution times on each problem have been used to establish the best predictor of performance. Outcomes of these analyses, in turn, have been used to infer which strategy generated the data. The same type of regression analyses were conducted in the present experiment but with a different purpose. The goal here was to establish whether performance of the present sample was comparable to that of children in previous studies that used chronometric data.

Six predictors were included in the regression analysis: the larger number, the smaller number, the difference between the larger and smaller number, the sum of the larger and smaller number, the prediction of the smaller-count model, and whether the smaller number exceeded 10. As discussed by Woods et al. (1975), the first five of these predictors correspond to alternative models of children's subtraction. For example, if children always started at the larger number and counted down the number of steps indicated by the smaller number, the size of the smaller number would be expected to be the best predictor. The predictor associated with the smaller-count model was the number of steps needed to count up or count down, whichever was smaller. Thus, the

| Table 1 |
| Strategy Use and Solution Times (in Seconds) of Each Strategy |

<table>
<thead>
<tr>
<th>Grade level</th>
<th>Retrieval %</th>
<th>Count down %</th>
<th>Count up %</th>
<th>Delete 10s %</th>
<th>Addition reference %</th>
<th>Guess-no response %</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>29</td>
<td>44</td>
<td>48</td>
<td>89</td>
<td>1</td>
<td>99</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>15</td>
<td>39</td>
<td>5</td>
<td>62</td>
<td>26</td>
<td>95</td>
</tr>
<tr>
<td>Overall</td>
<td>38</td>
<td>1.8</td>
<td>32</td>
<td>7.8</td>
<td>3</td>
<td>7.8</td>
<td>4</td>
</tr>
</tbody>
</table>

*Note* Overall solution times are medians of second and fourth graders' combined times.
predictor associated with the smaller-count strategy for both 12 - 9 and 12 - 3 was 3, whereas that for both 15 - 2 and 15 - 13 was 2. Because children erred on a relatively high percentage of trials (16%), because the errors were generated by the same strategies as the correct answers, and because removing trials with errors would seriously bias estimates of solution times on the most difficult problems, both correct and incorrect trials were used in computing the median solution time on each problem.

As in previous studies of subtraction by second through fourth graders, the smaller-count model proved to be the best predictor of the averaged solution times. The absolute percentage of variance in median solution times on each problem that was accounted for by the smaller-count model (52%) was also comparable to that in previous studies. A further parallel between the present results and previous ones was that no other predictor accounted for the solution time pattern nearly as well as the smaller-count model. The next best predictor of the solution times, the sum of the minuend and subtrahend, accounted for only 27% of the variance when it was entered into the regression equation first.

All of these results held true at both second- and fourth-grade levels. The smaller-count strategy was the best predictor of second graders' median solution times on each problem, accounting for 57% of the variance in the times. It was also the best predictor of fourth graders' median solution times on each problem, accounting for 38% of the variance. The next-best predictor in both cases was the sum, which accounted for 15% less variance in the second graders' times and 18% less variance in the fourth graders' times (42% and 20% of the variance, respectively) when it was entered into the regression equation first.

Analyses of times produced by each strategy. To test whether children actually used the strategies that they said they did, separate multiple regression analyses were conducted of those solution times classified as being produced by each strategy. For example, such an analysis was performed on the median solution time on each problem that was produced when children were classified as counting down from the larger number. To avoid unstable estimates of solution times, only problems for which the given strategy was classified as being used by at least 3 children were included in these analyses. This criterion allowed analysis of all 36 problems for retrieval, counting down, and addition reference; of 18 problems for guessing; of 17 problems for deleting 10s; and of 12 problems for counting up.

Analyses of the times classified as being generated by each strategy lent considerable convergent validity to the strategy assessments. Of particular importance were results for the two strategies, counting down and counting up, for which the nature of the strategy suggested what the best predictor of solution times should be. For trials on which children's overt behavior and self-reports led to their being classified as having counted down, the size of the smaller number, which indicates the number of downward counts, was by far the best predictor of performance. It accounted for 81% of the variance in solution times. For trials on which children's overt behavior and self-reports led to their being classified as having counted up, the difference between the minuend and the subtrahend, which corresponds to the number of upward counts needed to proceed from the smaller to the larger number, was the best predictor. It accounted for 38% of the variance. Neither of these variables was the best predictor of solution times for trials on which children reported using any of the other strategies. Thus, the combination of the immediately retrospective verbal reports and the videotaped records of ongoing problem-solving behavior allowed valid classification of children's strategies on each trial.

Strategy Use

Test of predictions of smaller-count model. A strength of Woods et al. 's (1975) smaller-count model is its unambiguous predictions about which strategy will be used on each problem. The model predicts that children will always count down on problems for which that approach requires fewer counts than counting up and that they will always count up on problems for which that requires fewer counts than counting down. The trial-by-trial assessments of strategy use in the present experiment allowed this prediction to be tested.

Table 1 demonstrates an extreme lack of fit between the strategy use predicted by the smaller-count model and the observed frequency of strategies. The counting-up strategy, which was predicted to be used on 50% of trials for this problem set, in fact was used on only 3%. Counting up was only the sixth most common strategy, less common not only than counting down but also than retrieval, addition reference, deleting 10s, and guessing or not responding (Table 1). Only 36% of children ever indicated that they counted up.

The predictions of the smaller-count model concerning when children would count-down also were radically inaccurate. Not only was the strategy used considerably less often than predicted (32% vs. 50%), it was used as much on the problems for which it was predicted never to be used (31%) as on the problems for which it was predicted always to be used (32%).

Overall, the smaller-count model correctly predicted children's strategies on only 19% of trials. Adherence was higher

Table 2: Use Predicted by Smaller-Count Model and Observed in Children's Performance on Counting Down and Counting Up Problems

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Type of problem</th>
<th>Counting down (e.g., 15 - 2)</th>
<th>Counting up (e.g., 15 - 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counting down</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% use predicted by model</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>% use observed</td>
<td>32</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Counting up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% use predicted by model</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>% use observed</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Note. Counting down problems are those for which counting down requires fewer counts than counting up (e.g., 15 - 2). Counting up problems are those for which counting up requires fewer counts than counting down (e.g., 15 - 13).
among second graders (25% of trials) than among fourth graders (13%) but was low in both grades. At the level of individual subjects, only 2 of the 74 children used the predicted type of counting on even one half of the trials; none of the 74 children used the predicted type of counting on as many as three-fourths of trials. Thus, despite the ability of the smaller-count model to predict the median solution time on each problem, literally no children consistently used the approach described by the model.

How can a strategy that is not used predict chronometric data? The results reported up to this point may seem paradoxical. The smaller-count model predicted median solution times on each problem quite well. However, the trial-by-trial strategy classifications indicated use of a much wider range of strategies than suggested by the smaller-count model. Furthermore, the counting up and counting down strategies were not assigned to problems in anything like the way predicted by the model. These failed predictions cannot be attributed to the strategy classifications' being invalid. Separately analyzing the solution times classified as being produced by each strategy yielded convergent validation for the strategy classifications; the predictors of solution times for each strategy differed, and the differences were intuitively reasonable (e.g., size of the smaller number was the best predictor of response times (RTs) on counting down trials; size of the difference was the best predictor on counting up trials). How, then, could the smaller-count model fit the averaged solution times as well as it did?

Figure 1 provides part of the answer. In this analysis, the 36 problems are divided into two groups of 18. The problems in the graph on the left are those for which the smaller count involves counting down. The problems in the graph on the right are those for which the smaller count involves counting up. The panel on the left indicates that solution times for problems for which the smaller count involved counting down followed the expected pattern. They increased with the size of the smaller number. In contrast, the panel on the right indicates that on problems for which the smaller count involved counting up, the solution times did not generally fit the predicted pattern. Only on problems where the size of the difference was 1 was the expected pattern apparent. That is, on problems such as 13 - 12, where the difference was 1, solution times were very short, as short as on problems such as 13 - 1 that involved counting down one number. However, on the other problems for which Woods et al. (1975) model predicted that children would count up, times were considerably longer and did not fit any clear pattern. (This lack of a discernible pattern in the right side panel can be seen clearly by putting one's finger over the data points for the problems with differences equal to 1.)

To test this interpretation, two separate regression analyses were conducted. One involved times on the 21 problems for

**Figure 1** Median solution times on problems for which smaller-count model predicts that children will count down (left panel) and on problems for which smaller-count model predicts that children will count up (right panel).
which the smaller-count model was believed to work reasonably well: the 18 problems for which children were hypothesized to count down and the 3 problems for which they were hypothesized to count up and for which the difference between the minuend and subtrahend was 1. The other analysis involved times on the 15 problems for which children were hypothesized to count up and for which the difference was greater than 1.

Within the analysis of times on the 21 problems for which the smaller-count model appeared to fit the data reasonably well, the model's predictions accounted for 64% of the variance in the median solution times. The next-best predictor, the size of the difference, accounted for only 17% In contrast, in the analysis of the times on the other 15 problems, the smaller-count model explained only 9% of the variance. The sum was a considerably better predictor (though still not an especially good one), accounting for 23% of the variance.

It is important to note that the 3 counting up problems for which the smaller-count model accurately predicted solution times (13 - 12, 14 - 13, 15 - 14) were the problems with the smallest differences of any problems presented in this set. Therefore, they were problems for which the Woods et al. smaller-count model yields a prediction of extremely fast solution times. The fact that times were, in fact, very fast on these problems greatly aided the fit of the smaller-count model to the overall data set.

Even on these three counting up problems, the fit of the smaller-count model was not due to consistent or even frequent use of counting up. The strategy classifications indicated that children used the counting up strategy on only 3% of trials on the three problems. Rather, the fast solution times on these problems were due to frequent use of retrieval. On all three problems, children used retrieval on 50% to 55% of trials. No other problem in the right hand graph had nearly as high a percentage of retrieval (Figure 2). In fact, among the entire set of 36 problems, the only other problems with such high levels of retrieval were the 9 problems with both minuend and subtrahend below 10. Retrieval in general led to much faster RTs than did other approaches. This was true for these three problems as well. On the 52% of trials on the three problems on which children were classified as using retrieval, their median RT was 1.4 s; on the other 48%, it was 5.0 s. Thus, frequent use of retrieval on the three problems, rather than counting up, apparently led to the fast solution times. These fast solution times, together with the reasonably good fit of the model to times on the “counting down” problems, led to the ability of the smaller-count model to predict solution times for the data set as a whole.

Discussion

As in previous studies of children's subtraction, the smaller-count model accurately predicted average solution times on each problem. However, a variety of other data indicated that the cognitive processes hypothesized by the smaller-count

*Figure 2  Percentage of retrieval on problems for which smaller-count model predicts that children will count down (left panel) and on problems for which smaller-count model predicts that children will count up (right panel)
model bore little resemblance to those that generated the solution times. Children appeared to use a wide variety of subtraction strategies, not just counting up and counting down. They counted up on fewer than 5% of trials, not the 50% that the model predicted. They counted down as often on problems on which, according to the model, they should never count down as on problems on which, according to the model, they should always count down.

The findings have implications not only for children's arithmetic but for a wide range of other cognitive activities. These implications concern the role of mental chronometry in inferring processing strategies, how strategy use can be assessed in situations in which people use multiple strategies, and when people are most likely to use multiple strategies.

**Uses and Abuses of Mental Chronometry**

The present experiment illustrated two characteristics of chronometric methods, one negative and one positive. The negative characteristic was probably more evident. Standard chronometric analyses are based on the assumption that the same strategy is used on every trial. If this assumption is violated, such methods often will yield misleading results. This holds true even when the fit between the model and the averaged data is good. Siegler (1987a) illustrated how an addition strategy that was consistent with a large body of chronometric data accurately described what subjects were doing on only about one third of trials. The present case of second and fourth graders' subtraction was yet worse. The smaller-count model accurately predicted which strategy children would use on only one fifth of trials. No child adhered to it consistently.

The reasons for the apparent success of these models in earlier studies were different in subtraction than in addition. In the case of addition, the min model fit the chronometric data averaged across strategies because the min strategy was one of the two strategies that children most often used, and it contributed much more variance in solution times than did the other frequently used strategy, retrieval (Siegler, 1987a). In contrast, the smaller-count model fit the present data on subtraction because it accurately described times for two subsets of the data. The model fit the data on problems for which it predicted that children would count down, because children actually did count down quite often on these problems. However, its fit to another subset of the data, problems on which the difference between the two numbers was 1, was not due to counting up being a frequent strategy on these problems. It was used on only 3% of trials on them. Rather, the accurate predictions were due to frequent use of retrieval on these problems, which led to the very fast solution times predicted by the smaller-count model for them. On the 21 problems in these two subsets of the data, the smaller-count model accounted for more than 60% of the variance in solution times; on the remaining 15 problems, the model accounted for less than 10% of the variance.

But there is also a positive side to the story. Even when people use multiple strategies, chronometric methods can make crucial contributions. These contributions come about when the solution times are classified according to the strategy that generated them and then separately analyzed. When used in this way, chronometric analyses can be invaluable in validating or disconfirming strategy assessments derived from other kinds of data.

In the case of children's subtraction, although chronometric methods provided the earlier evidence that seemed to support the smaller-count model, in the present study they provided essential evidence for its incorrectness. The chronometric analyses that indicated that the size of the smaller number was the best predictor of solution times when children were classified as counting down and that the size of the difference was the best predictor when they were classified as counting up were critical sources of support for the validity of the strategy classifications. Without these analyses, the validity of children's verbal reports would have been uncertain.

What lessons can be derived from these positive and negative contributions of chronometric methods? One lesson is that when people use diverse strategies, chronometric analyses of data averaged over the strategies are a problematic means for inferring strategy use. Analysis of other kinds of data seems essential for accurate identification of strategies. If such data on strategy use can be obtained, though, chronometric analyses again become invaluable, because they can yield converging (or nonconverging) evidence for the strategy assessments.

**Assessing Use of Multiple Strategies**

Ericsson and Simon (1984) concluded that people can often accurately describe their processing when they report immediately after the processing episode and when the processing is not extremely brief in duration. The present study met these criteria and provided evidence consistent with Ericsson and Simon's conclusion.

The very success of indirect indexes of cognitive activity, such as solution times, may have led to an overemphasis on their use and to excessive skepticism about verbal reports as data. As has frequently been noted (e.g., Brainerd, 1973; Nisbett & Wilson, 1977), self-reports and verbal explanations can yield misleading impressions of people's thinking. The present study, however, demonstrates that analyses based on hard data, such as solution times, can be equally misleading. Understanding of people's strategies in many situations may benefit greatly from obtaining immediately retrospective verbal reports and using them to better understand the meaning of solution times, errors, and other "hard" data.

In many situations, it is impossible to obtain meaningful verbal reports. In other situations, verbal reports are invalid. What can be done to assess strategy use in these situations? The present findings illustrated one source of relevant data: the extent to which subsets of the data fit the model. If the fit is uneven, as Figure 1 illustrates it is in the present case, there is reason to suspect models that postulate a single strategy. In particular, the fact that the model fit the data for problems on which counting down was expected, but not most of those on which counting up was predicted, provided reason to suspect that counting up might not be used very often.

Examining distributions of times within particular sets of problems, rather than just the average values, also can be
useful for suggesting that multiple strategies are being used.
For example, on the problems 13–12, 14–13, and 15–14, 59% of solution times were below 3 s, 15% of times were
between 3 and 6 s, and 25% of times were above 6 s. This
type of nonnormal distribution suggests that the data may
have been generated by more than one strategy. Obtaining
verbal reports during the course of problem solving, examin-
ing error patterns, and identifying consistent clusters in recall
are other methods that can prove useful for identifying when
people are using multiple strategies (Pellegrino & Ingram,
1979; Siegler & Shrager, 1984; Staszewski, 1987; VanLehn,
1983).

When Do People Use Multiple Strategies?

People use diverse strategies not only in arithmetic but in
a wide variety of educationally relevant domains. For exam-
ple, in spelling, children sometimes rely on rules that connect
sounds and symbols, sometimes draw analogies to other words
whose spelling is better known, sometimes form mental im-
age of how the word looked in the past, sometimes rely on
word-specific heuristics (accommodation has two cs and two
ms), sometimes look up words in dictionaries, and sometimes
retrieve spellings from memory (Ehri, 1980; Marsh, Fried-
man, Welch, & Desberg, 1980; Siegler, 1986). In telling time
on an analog clock, children sometimes count forward or
backward by 5s or 1s from the hour; sometimes count forward
or backward from landmarks such as quarter hours, half
hours, and 5-min marks; and sometimes retrieve the time
from memory (Siegler & McGilly, 1989). Reading (Just &
Carpenter, 1987), writing (Bereiter & Scardamalia, 1987),
and mathematics beyond arithmetic (Mayer, 1982) also elicit
use of diverse strategies.

The degree of knowledge about the task seems to be one
important determinant of when people use multiple strategies.
When first acquiring knowledge about a domain, learners
may typically know only a single strategy. When learners have
had a great deal of experience with particular problems, they
often become able to solve them through a single, very effi-
cient approach. For example, adults do not typically need to
use any approach other than retrieval to tell time or to solve
simple arithmetic problems. It is in the intermediate phase,
when several strategies have been acquired but problems
cannot all be solved very efficiently, that use of diverse strat-
egies seems most prominent.

Implications for an Understanding of Arithmetic

The present results indicate a number of characteristics of
children’s arithmetic that any adequate model needs to reflect.
Such a model needs to depict children as using a wide variety
of strategies while they are acquiring arithmetic skills. This
variability of strategy use must be present within, as well as
between, individuals. It also must be present within, as well
as between, problems; children do not simply use one strategy
on one type of problem and a different strategy on other types
of problems. The model further needs to explain how with
age, the speed and accuracy with which children execute
different strategies increases. It also needs to account for their
progressive movement toward more frequent use of the faster
strategies, such as retrieval and addition reference, and for
their decreasing use of the slower strategies, such as counting
down and guessing. Finally, an adequate model needs to
account for children’s ability at all ages to choose among
strategies so as to use each strategy most often on problems
for which its advantages in terms of speed and accuracy are
greatest, relative to those of other strategies. Such a model
would increase understanding not only of acquisition of arithmetic
skills but of many other cognitive activities as well.

References

composition Hillsdale, NJ: Erlbaum
Brainerd, C. J. (1973) Judgments and explanations as criteria for the
presence of cognitive structures Psychological Bulletin 79: 172–
179
Carpenter, T. P. (1986) Conceptual knowledge as a foundation for pro-
cedural knowledge In J. Hiebert (Ed.), Conceptual and proc-
cedural knowledge: The case of mathematics (pp 113–132) Hillsdale,
NJ: Erlbaum
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
and subtraction concepts in grades one through three Journal for
Research in Mathematics Education 15: 179–202
Frith (Ed.), Cognitive processes in spelling (pp 311–338) New
York: Academic Press
reports as data Cambridge, MA: MIT Press.
Fuson, K. C. (1982) An analysis of the counting-on solution proce-
dure 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
Research in Mathematics Education 15: 214–223
Cognitive addition: Comparison of learning disabled and academ-
ically normal elementary school children Cognitive Development.
2: 249–269
practice of basic addition facts: Strategy changes in learning dis-
able students Cognition and Instruction 5: 223–265
Groen, G. J., & Parkman, J. M. (1972) A chronometric analysis of
simple addition Psychological Review 79: 329–343
Groen, G. J., & Poli, M. (1973) Subtraction and the solution of open
sentence problems Journal of Experimental Child Psychology. 16:
292–302
addition algorithms? Journal of Education Psychology 69: 645–
652
and second-grade children Journal for Research in Mathem-
atics Education. 12: 95–106
Journal of Genetic Psychology 79: 3–28


In I. Levin & D. Zakay (Eds.), *Psychological time: A life span perspective* (pp. 185–218) New York: North Holland.


Received August 4, 1988
Revision received December 26, 1988
Accepted January 30, 1989