Individual Differences in Strategy Choices: Good Students, Not-So-Good Students, and Perfectionists

Robert S. Siegler
Carnegie-Mellon University

SIEGLER, ROBERT S. Individual Differences in Strategy Choices: Good Students, Not-So-Good Students, and Perfectionists. Child Development, 1988, 59, 833–851. Consistent individual differences were found in first graders’ strategy choices in addition, subtraction, and reading (word identification). Differences were present along 2 dimensions: knowledge of problems and stringency of thresholds for stating retrieved answers. Cluster analyses indicated that children could be classified into 3 groups: good students, not-so-good students, and perfectionists. Perfectionists were children who had good knowledge of problems and set very high thresholds for stating retrieved answers, good students also had good knowledge of problems but set lower thresholds, and not-so-good students had less good knowledge of problems and set low thresholds. Differences among the 3 groups were evident on measures not included in the cluster analysis as well as measures that were. Further, the groups differed in standardized achievement test performance 4 months after the experiment in ways consistent with the experimental analysis. The pattern of individual differences was similar in 2 experiments with different samples of children and problems and different methods for assessing strategy use. The results illustrated how detailed cognitive models can contribute to understanding of individual differences.

The purpose of this article is to examine individual differences in children’s strategy choices. In particular, the research focuses on consistencies in 6-year-olds’ strategy choices on three tasks: addition, subtraction, and word identification (reading). The issues of greatest interest are what types of consistent individual differences exist in children’s strategy choices; how such individual differences, if present, can be interpreted within the framework of a previously formulated model of children’s strategy choice procedures; and how individual differences in children’s strategy choices in the experimental situation relate to their standardized test performance.

If individual children knew only a single strategy for performing a given task, there would be no need to determine how they choose among multiple strategies, much less to examine individual differences in how they choose among the strategies. Recent studies, however, indicate that individual children and adults often use multiple strategies for solving a given problem. They do so in such diverse areas as referential communications (Kahan & Richards, 1986), series completion (LeFevre & Bisanz, 1986), addition and subtraction (Siegler & Shrager, 1984), question answering (Reder, 1987), and causal reasoning (Shultz, Fischer, Pratt, & Rulf, 1986).

The fact that people use diverse strategies is not a mere idiosyncrasy of human cognition. Strategies differ in their accuracy, in the amounts of time needed for execution, in their memory demands, and in the range of problems to which they apply. Wise choices of strategies allow people to meet situational demands and to overcome limited knowledge and processing resources.

Even young children often exhibit considerable skill in choosing strategies. This skill is evident in their choices of whether to state a retrieved answer or to use a backup

This research was supported in part by grant HD-19011 from the National Institutes of Health and in part by grant 83-0050 from the National Institute of Education. Thanks are due to Miss Hawkins, Miss Ferry, and the teachers and students of University Park Elementary of Monroeville, PA, who made the research possible, as well as to Denise Wilk and Cathy Dennler, who conducted the experiment and analyzed the data. Special thanks to Jamie Campbell, who suggested many useful ideas about the research. Requests for reprints should be sent to Robert S. Siegler, Psychology Department, Carnegie-Mellon University, Pittsburgh, PA 15213.

[Child Development, 1988, 59, 833–851. © 1988 by the Society for Research in Child Development, Inc. All rights reserved. 0009-3920/88/5904-0003$01.00]
strategy (a strategy other than retrieval). Examples of backup strategies include counting up from one to add, counting down from the larger number to subtract, sounding out words to read, and looking up a word's spelling in a dictionary. Both retrieval and use of backup strategies have clear, though different, advantages. Retrieval can be executed much faster; backup strategies often yield high accuracy rates on problems where retrieval cannot. Ideally, children would use retrieval when it could be executed accurately and would use backup strategies where they were necessary for accurate performance. In fact, even young children's strategy choices consistently follow such a pattern in many different domains (Siegler, 1986).

How do children reach such adaptive strategy choices? Siegler and Shrager (1984) developed the distributions-of-associations model to describe the strategy choice process. The model is designed to account for which strategies children use, the relative solution times of the strategies, the relative difficulty of problems, how strategies are chosen at any one time, and how strategy choices, solution times, and error rates change over time. It has been subjected to considerable empirical testing in addition, subtraction, and multiplication and has been formalized as a running computer simulation that both performs and learns in those areas (Siegler, 1987a, 1987b; Siegler & Shrager, 1984). The model's complexity precludes a full description of it here; extensive descriptions of it are available in the sources cited previously. However, the model's main mechanism for choosing between stating a retrieved answer and using a backup strategy is both relatively simple and critical to the present analysis of individual differences; therefore, I will describe it in some detail.

This strategy-choice mechanism involves two interacting parts: a representation of knowledge about particular problems, and a retrieval process that operates on the representation to produce performance. The representation involves associations of varying strengths between each problem and answers, both correct and incorrect, on that problem. For example, in Figure 1, the answer 6 is connected to the problem "3 + 5" with a strength of .19, the answer 7 is connected to "3 + 5" with a strength of .17, and so on. Representations of different problems can be thought of as varying along a dimension of peakedness. In Figure 1, the representation of "2 + 1" is a peaked distribution because the preponderance of associative strength is concentrated in a single answer (the peak of the distribution). The representation of "3 + 5," in contrast, is a flat distribution because associative strength is distributed among a number of answers, with no one of them constituting a strong peak.

The process operates on this representation in the following way. First, the child sets a confidence criterion. This confidence criterion is a threshold that must be exceeded by the associative strength of a retrieved answer for that answer to be stated. It can assume any of a range of numerical values. Once this threshold is set, the child retrieves an answer.
The probability of any given answer's being retrieved on a particular retrieval effort is proportional to the associative strength of that answer relative to the associative strengths of all answers to the problem. Thus, because in the Figure 1 example the associative strength connecting "2 + 1" and "3" is .82, and because the total associative strength connecting "2 + 1" with all answers is 1.00, the probability of retrieving "3" as the answer to "2 + 1" would be .82. If the associative strength of whatever answer is retrieved exceeds the confidence criterion, the child states that answer. Otherwise, the child may either again retrieve an answer and see if it exceeds the confidence criterion or abandon efforts to retrieve and instead use a backup strategy to solve the problem.

Within this model, both the peakedness of distributions of associations and the stringency of confidence criteria influence performance. First consider the role of the peakedness of distributions of associations. The more peaked a given distribution, the more often that retrieval, rather than a backup strategy, will be used (because the greater the concentration of associative strength in one answer, the higher the probability that the most strongly associated answer will be retrieved, and the higher the probability that that answer's associative strength will exceed the confidence criterion and thus allow the retrieved answer to be stated). Because the answer with the greatest associative strength ordinarily is the correct answer, the greater the concentration of associative strength in that answer, the more probable that the retrieved answer will be correct. Finally, the greater the concentration of associative strength in one answer, the more probable that solution times will be short because children will be likely to retrieve stable answers early in the retrieval process.

This analysis implies that percent backup strategy use on each problem should correlate highly with speed and accuracy on that problem; all three variables are functions of the peakedness of the problem's distribution of associations. Consistent with this prediction, percent use of backup strategies on each problem has correlated at least $r = .80$ with speed and accuracy on that problem in addition, subtraction, multiplication, and time-telling (Siegler, 1986; Siegler & McGilly, in press; Siegler & Shrager, 1984). A further, quite nonintuitive prediction is that percent use of backup strategies on each problem will correlate more highly with speed and accuracy on retrieval trials on that problem than with speed and accuracy on backup strategy trials on the problem. Speed and accuracy on retrieval trials on the problem should be determined by the peakedness of the distribution of associations, just as percent use of backup strategies is. In contrast, speed and accuracy on backup strategy trials should be determined by the specific factors contributing to difficulty of executing the backup strategies on the problem. Consistent with this prediction, percent backup strategy use on each problem has correlated significantly more highly with speed and accuracy on retrieval trials on the problem than with speed and accuracy on backup strategy trials on the problem in all four of the above-cited domains. Thus, peakedness of distributions seems to influence when backup strategies are used.

The role of the confidence criterion has not been studied empirically. However, systematic manipulations within the computer simulation of addition, as well as simple logic, indicate that the higher the confidence criterion, the fewer the associative strengths that exceed it. Thus, other things equal, high confidence criteria lead to frequent use of backup strategies and low confidence criteria to frequent use of retrieval.

Implications for Analyzing Individual Differences

Individual differences in strategy choices are of interest for both empirical and theoretical reasons. The basic empirical issues are whether individual children show consistent patterns of strategy choices across different tasks, what the nature of the differences are, and how differences in strategy choices relate to standardized test performance. The basic theoretical issue is whether individual differences in strategy choices can profitably be analyzed in terms of the dimensions of variation that are central within the present model, the peakedness of distributions of associations, and the stringency of confidence criteria.

In the present experiments, children's strategy choices were examined on three tasks: addition, subtraction, and word identification. These tasks were chosen both because the model had been found to apply to all of them and because they allowed a number of potentially interesting comparisons regarding individual differences. Several considerations suggested that individual differences in strategy choices in addition and subtraction would be the most closely related. Addition and subtraction are both numerical tasks. Children use very similar strategies on them, for example, adding by counting up from the larger
number in the problem and subtracting by counting down from the larger number in the problem. Addition and subtraction also are usually taught by the same teacher using the same textbook; this may contribute motivational similarities (e.g., a boy wants very much to do well in math because he likes the subject and the teacher, and therefore sets high confidence criteria in both addition and subtraction).

Word identification does not share any of these qualities with addition and subtraction. However, it does appear to share with them the same basic strategy choice procedure. Firth (1972) and Jorm and Share (1983) have suggested that from early in the learning-to-read process children use visually based retrieval when that approach can yield correct pronunciations, and backup strategies such as phonological recoding when such backup strategies are the only likely way to generate a correct answer. Siegler (in preparation) found that the pattern of use of retrieval and backup strategies in word identification was much like that in arithmetic. Thus, although arithmetic and word identification backup strategies differ substantially, and although specific motivational factors provided by teachers and textbooks may vary, similar strategy choice procedures, along with general ability and motivational factors, may lead to similarities among strategy choices on all three tasks.

Beyond revealing overall relations among the tasks, the experiment also provided an opportunity to determine whether individual children's performance fell into consistent group patterns more general than the performance of the individual child but more specific than the central tendency of the sample as a whole. The goal was to identify groups of children whose performance was alike in numerous respects and whose performance consistently differed from that of children in the other groups.

In addition to identifying individual differences in strategy choices, a further goal of the study was to examine the usefulness of the strategy choice model for interpreting whatever individual differences emerged. The model had previously been found useful in analyzing differences in strategy choices on different problems; the new question was whether it would also be useful in analyzing differences in strategy choices of different individuals.

The model's basic assumption regarding individual differences in strategy choices is that they arise through parametric variation within a single basic procedure, the one depicted by the model. That is, different children are viewed as using the same procedure for choosing among strategies; differences in the strategy choices that they make are viewed as being due to differing values of parameters that influence strategy choices. The two main parameters of the model that influence strategy choices are the peakedness of distributions of associations and the stringency of confidence criteria. Individual differences in peakedness of distributions would lead children with peaked distributions to retrieve more often, more accurately, and more quickly than children with flat distributions. Individual differences in confidence criteria would lead children who set high criteria to retrieve less often, and to use backup strategies more often, than children who set lower criteria.

Experiment 1

METHOD

Participants.—The children were 21 first-grade boys and 15 first-grade girls attending a middle-class suburban public school in the Pittsburgh area. Their mean age at the time of testing was 81 months (SD = 3.5 months). The experimenter was a 30-year-old female research assistant.

Problems.—The addition problems were 14 items with larger addends ranging from 2 to 15, smaller addends from 1 to 6, and sums from 3 to 18. The subtraction problems were the 14 exact inverse problems from those presented in addition. The reading items were 50 words ranging in difficulty from two-letter-words ("in") to eight-letter-words ("sandwich"). Each word was printed in lowercase letters on a 4 × 6-inch index card. The words and arithmetic problems were sampled from those in the children's textbooks (Addison-Wesley Series, 1985, in arithmetic; Scott-Foresman Series, 1976, in reading). Approximately 70% of the words and problems had appeared in lessons that the children had already completed before the time of testing; the remaining items were in lessons that the class had not yet encountered.

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 each session, children were told that on that day they would be adding, subtracting, or reading. (A child did only one task on a given day.) Equal numbers of children were presented each of the three tasks in
TABLE 1
SPEED AND ACCURACY ON EACH TASK: EXPERIMENT 1

<table>
<thead>
<tr>
<th>Task</th>
<th>% Retrieval</th>
<th>% Correct</th>
<th>Median RT</th>
<th>% Correct, Retrieval Trials</th>
<th>% Correct, Backup Strategy Trials</th>
<th>Median RT, Retrieval Trials</th>
<th>Median RT, Backup Strategy Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>61</td>
<td>76</td>
<td>9.9</td>
<td>82</td>
<td>75</td>
<td>4.2</td>
<td>19.4</td>
</tr>
<tr>
<td>Subtraction</td>
<td>54</td>
<td>64</td>
<td>11.6</td>
<td>69</td>
<td>61</td>
<td>6.1</td>
<td>22.4</td>
</tr>
<tr>
<td>Reading</td>
<td>74</td>
<td>50</td>
<td>2.5</td>
<td>89</td>
<td>49</td>
<td>1.8</td>
<td>20.9</td>
</tr>
</tbody>
</table>

Each of the six possible orders. The three tasks were presented on three consecutive school days; these usually were literally consecutive days, though at times they were separated by a weekend.

The instructions emphasized that the child could use any approach to get the right answer; all that was important was that the child try as hard as possible. For example, the instructions given on the reading task were, "I'm going to show you some reading words. I want you to take your time and try to read the words that I show you. You can do anything you want to figure out what word it is: sound it out, just know what it is, recognize a little word inside the big word, or whatever. Just do the best that you can." The instructions for the addition and subtraction tasks were similar; the exact wording is given in Siegler and Shrage (1984).

Each child's behavior was video-recorded using a Sony SLO-323 videocassette recorder and a Sony 3260 camera. Solution times were recorded through use of a Vicon X240 digitizer. This apparatus prints digital times, accurate to \( \frac{1}{10} \) of a second, across the bottom of the taped scene. To supplement the videocassettes, the experimenter made notes about the child's observable behavior on each problem where such behavior was detected.

The standardized test that was related to the children's experimental performance was the Metropolitan Achievement Test (1985 Revision, Form R). This test included six scores that seemed relevant to the present experiment: total mathematics, mathematics computation, mathematics problem solving, total reading, word recognition, and reading comprehension. Children were given this test at the end of the school year, 4 months after the experimental sessions.

RESULTS AND DISCUSSION

The results are divided into three parts. The first presents an overview of children's strategy use, accuracy, and speed on each task. The second examines relations of individual performance across the pairs of tasks. The third examines consistencies of individual performance across all three tasks.

Performance on Each Task
Table 1 provides an overview of performance on each task. For purposes of comparability across tasks, the different strategies that children used were reduced to two categories: retrieval and backup strategies. Children were classified as using backup strategies when they produced overt behavior (visible or audible behavior) between presentation of the problem and statement of the answer. The most common backup strategies were counting up from one or from the larger number in addition, counting down from the larger number in subtraction, and sounding out in reading. Classification of whether a child used retrieval or a backup strategy was very reliable; the classifications of two raters who independently watched the videotapes agreed on 100% of trials.

As can be seen in Table 1, performance was roughly comparable across the three tasks on most measures. Percent use of retrieval ranged from 54% to 74% on the three tasks; percent correct ranged from 64% to 80%. The only substantial difference was on solution times, where median times were much faster on the reading task (2.5 sec/word) than on the addition and subtraction tasks (about 10 sec/problem).

Tests of the model's predictions.— Performance within each of the tasks fit the model's predictions quite well. Because data on the fit of the present strategy choice model to word identification have not previously been reported, this task will be used to illustrate how the model fit the data (for data on the models' fit to addition and subtraction in this age range, see Siegler, 1987a; Siegler & Robinson, 1982; Siegler & Shrage, 1984).
### TABLE 2
CORRELATIONS OF INDIVIDUAL CHILDREN'S PERFORMANCE IN ADDITION, SUBTRACTION, AND READING: EXPERIMENT 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Addition-Subtraction</th>
<th>Addition-Reading</th>
<th>Subtraction-Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>% retrieval</td>
<td>.72</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>% correct</td>
<td>.60</td>
<td>.36</td>
<td>.43</td>
</tr>
<tr>
<td>Mean RT</td>
<td>.98</td>
<td>-.09</td>
<td>-.03</td>
</tr>
<tr>
<td>% correct on retrieval trials</td>
<td>.48</td>
<td>.59</td>
<td>.61</td>
</tr>
<tr>
<td>Mean RT on retrieval trials</td>
<td>.70</td>
<td>.31</td>
<td>.27</td>
</tr>
<tr>
<td>% correct on backup strategy trials</td>
<td>.63</td>
<td>.21</td>
<td>.31</td>
</tr>
<tr>
<td>Mean RT on backup strategy trials</td>
<td>.88</td>
<td>-.04</td>
<td>.03</td>
</tr>
</tbody>
</table>

**NOTE.**—All correlations are Pearson product-moment r's. With df = 34, r's > .33 are significant for p < .05.

As in previous studies, there were two main tests of the model's fit to performance. One was the prediction of high correlations between percent backup strategy use, percent errors, and median solution time on each problem. The other was the prediction that the sources of these correlations would be primarily the patterns of errors and solution times on retrieval trials.

The reading task data were in accord with both predictions for both errors and solution times. The correlation between percent backup strategy use on each word and percent errors on that word was r = .83. The source of these correlations was primarily errors and solution times on retrieval trials. Percent errors on retrieval trials on each word correlated r = .67 with percent backup strategy use on the word. This was a significantly stronger relation than the r = .39 between percent errors on backup strategy trials on each word and percent backup strategy use on that word, dependent measures, t(38) = 2.29, p < .05. The difference between the correlations was also present when length of the word, which was expected to be a major determinant of the first graders' difficulty of sounding out, was partialed from both correlations, r = .54 versus r = .17, t(38) = 2.36, p < .05.

Similar differences separated the correlations involving solution times on the two types of trials. Percent backup strategy use on each word correlated r = .83 with median solution time on the word. The correlation between median solution times on retrieval trials on each word and percent backup strategy use on the word was higher than that between median solution times on backup strategy trials and percent backup strategy use on the word, r = .65 versus r = .03, t(38) = 3.78, p < .001. Again, the difference was also present when word length was partialed out, r = .52 versus r = .01, t(38) = 2.86, p < .01. Thus, the relations hypothesized by the strategy choice model proved to be consistently present on measures of accuracy and speed.

**Relations of Addition, Subtraction, and Word Identification Performance**

The main focus of the study was on consistencies in individual children's performance across tasks. First consider relations of performance on the pairs of tasks. As shown in Table 2, correlations between each child's addition and subtraction performance were significant and substantial, both when performance on all trials was considered together and when performance on retrieval and on backup strategy trials was considered separately. All seven of the correlations were significant.

Relations between reading and the two arithmetic tasks were less consistent. On some measures, there were insignificant relations between which children did best on the word identification and on the arithmetic tasks; on other measures there were weak but significant relations; on one measure, the relations between word identification performance and performance on both addition and subtraction were stronger than the relation between addition and subtraction performance. In general, the pattern indicated that individual differences in addition and subtraction were closely related and that individual differences in word identification were

---

1 In keeping with the convention established in previous studies, the only items for which the correlations were computed were those on which both strategies were used at least three times. This standard was met on 41 of the 50 reading items, which resulted in the degrees of freedom reported.
somewhat related to those on the two arithmetic tasks.

Good Students, Not-So-Good Students, and Perfectionists

Perhaps the most intriguing of the data analyses was a cluster analysis, performed to indicate whether different children's performance on the three tasks fell into characteristic patterns. The input to the cluster analysis was nine data points for each child: three measures (percent retrieval, percent correct on retrieval trials, and percent correct on backup strategy trials) on each of the three tasks (addition, subtraction, and word identification). The clustering algorithm (PKM) is a nonhierarchical method that establishes a fixed number of homogeneous groups of cases using Euclidean distances (BMDP, 1981). Once the clustering algorithm divided children into groups, ANOVAs were used to determine the measures on which significant differences among the groups were present. Finally, the sources of differences among groups on each measure were probed further through use of Tukey honestly significant differences (HSD) post-hoc comparisons. Due to there being unequal numbers of children in the groups identified by the clustering algorithm, harmonic means were used to estimate "n" within the post-hoc comparisons (Hays, 1981). For all statistical tests, differences labeled as significant were significant at least at the .05 level.

Initially, the results of the cluster algorithm's two- and three-group solutions were compared. The three-group solution provided considerably better discrimination on the nine measures. Separate ANOVAs on each variable indicated that the two-group solution produced significant differences on only four of the nine measures; the three-group solution produced significant differences on eight of the nine (all but percent retrieval on the word identification task). Therefore, the three-group solution was the focus of further analyses.

To provide convergent validation for the three groups identified by the cluster analysis, ANOVAs were used to contrast the groups on six measures that were not used as part of the original input. These were median solution times for each child on retrieval trials on the three tasks and median solution times for each child on backup strategy trials on the three tasks. Performance of the three groups differed significantly on all four measures on the arithmetic tasks. The ANOVAs involving the two measures of reading performance did not reach significance, though the pattern of means was similar (Fig. 2). ANOVAs of the three groups' performance over all trials also revealed significant differences in percent correct on all three tasks and in median solution times on the addition and subtraction tasks.

Most important, each group's performance was readily interpretable in straightforward terms. For reasons that should become evident, I dubbed the three groups the "good students," the "not-so-good students," and the "perfectionists." The three groups included 12, 9, and 15 students, respectively. Differences between the good and the not-so-good students were present along all of the dimensions that might be expected from the names. Tukey HSD post-hoc comparisons indicated that the good students were correct significantly more often than the not-so-good students on retrieval trials on all three tasks, that they were correct significantly more often on backup strategy trials in addition and subtraction, that they were significantly faster on retrieval trials in addition and subtraction, and that they were significantly faster on backup strategy trials in addition and subtraction. As shown in Figure 2, differences on the other six measures, although non-significant, also consistently favored the good students.

The relation of the performance of the perfectionists to that of children in the other two groups was more complex. The perfectionists broke the usual relation among strategy use, errors, and solution times. The HSD tests indicated that relative to the not-so-good students the perfectionists were significantly more accurate on retrieval trials on all three tasks and on backup strategy trials in addition and subtraction. Despite this greater speed and accuracy, however, the perfectionists used retrieval significantly less often on the addition and subtraction tasks, and no more often on word identification (Fig. 2).

Similarly, the perfectionists were at least as fast and accurate as the good students, yet used retrieval much less often. HSD tests indicated that the two groups differed significantly on only one of the 12 measures of

* Special thanks to Jamie Campbell, whose idea it was to conduct this analysis.
speed and accuracy; perfectionists had a higher percent correct retrieval in subtraction. In terms of the direction of differences, the perfectionists' speed and accuracy were higher than those of the good students on nine of the 12 speed and accuracy measures (Fig. 2). However, the perfectionists were significantly less likely than the good students to use retrieval on both addition and subtraction and no more likely to use it in word identification.

The contrast between the perfectionists' relatively rare use of retrieval, yet excellent performance when they did state retrieved answers, can be seen especially vividly in comparing their percent correct with their percent use of retrieval. On the addition and subtraction tasks (the tasks on which the three groups performed most differently), the perfectionists were correct on 98% of retrieval trials, the good students on 80%, and the not-so-good students on 42%. However, despite their facility with retrieval, the perfectionists used retrieval on only 42% of trials, versus 80% for the good students and 74% for the not-so-good ones.

One possible explanation for the perfectionists' superior speed and accuracy on retrieval trials was that it was entirely due to their using retrieval on problems where it was easier to retrieve the correct answer. As documented above and in previous studies, children in general use retrieval most often on problems where it is easiest to retrieve correctly. The perfectionists, who used retrieval on roughly half of trials, could limit their use of retrieval to problems where all children could retrieve accurately. The good and not-so-good students, who used retrieval on roughly three-quarters of trials, could not be so selective. Thus, the perfectionists' superior accuracy on retrieval trials might be attributable totally to their only using retrieval on problems where retrieval accuracy was relatively high for everyone.

To test this possibility, an effort was made to determine whether there were differences in the three groups' percent correct retrieval that could not be explained on the basis of the problems on which children in each group used retrieval. This required calculating for each group the percent correct re-
retrieval that would be expected on the basis of that group's frequency of retrieval on each problem and the total sample's percent correct on the corresponding problem. Following this, differences in the observed percent correct retrieval of the three groups could be compared to differences in these expected percent correct retrievals.

The calculation of expected percent correct retrieval involved multiplying the percent use of retrieval on each problem of children in a particular group by the total sample's percent correct on retrieval trials on that problem, and then summing across all problems for that group. Dividing this sum by the group's percent use of retrieval summed across all problems yields an expected percent correct retrieval for the group (expected on the basis of the difficulty of items on which children in the group used retrieval). Thus, the formula for computing the expected percent correct retrieval for the trials on which children in group $j$ used retrieval was:

$$
\frac{\sum_{i=1}^{l} R_{ij} D_i}{\sum_{i} R_{ij}},
$$

where $l =$ number of problems, $R_{ij} =$ percent retrieval on problem $i$ by children in group $j$, and $D_i =$ percent correct retrieval on problem $i$ across all groups.

Results of this analysis indicated that the problems on which retrieval was used by children in each group were part of the explanation for the perfectionists' superior performance on retrieval trials, but not the whole story. The expected percent correct retrieval for problems on which the perfectionists used retrieval was somewhat higher than that for problems on which children in the other two groups did so. Averaged across the three tasks, the expected percent correct retrieval for the problems on which the perfectionists, good students, and not-so-good students used retrieval was 83%, 75%, and 74%, respectively. However, the three groups' observed percent correct retrieval, 97%, 80%, and 53%, showed greater differences than the expected percent correct retrieval would suggest. Subtracting the expected from the observed value for each group indicated that the perfectionists retrieved correctly on 14% more trials than their expected percent correct retrieval, that the good students retrieved correctly on 5% more trials than expected, and that the not-so-good students retrieved correctly on 21% fewer trials than expected. Thus, after correcting for expected percent correct retrieval of problems on which children in the three groups used retrieval, the perfectionists retrieved correctly on somewhat more trials than the good students (9%), and both groups retrieved correctly on considerably more trials than the not-so-good students (35% and 26%, respectively).

Sex differences.—Given the performance of the three groups and popular stereotypes concerning young children performing academic tasks, it might have been expected that most of the perfectionists would be girls and most of the not-so-good students would be boys. This was not the case, though. Percentages of boys and girls in the three groups were similar; 67% of the perfectionists were boys, as were 50% of the good students and 60% of the not-so-good students, $\chi^2(2) < 1$.

Achievement test performance.—On average, the children performed somewhat above the national average on the Metropolitan Achievement Test. Their mean total mathematics score was at the seventy-second percentile and their mean total reading score was at the seventy-fourth percentile.

Differences on the achievement test between perfectionists and good students on the one hand and not-so-good students on the other paralleled those that had been found in the experiment. As shown in Table 3, the perfectionists and the good students both scored consistently higher than the not-so-good students. The perfectionists' average subtest scores ranged from the seventy-sixth to the eighty-sixth percentile, the good students' average scores ranged from the sixty-eighth to the eighty-fourth percentile, and the not-so-good students' average scores ranged from the twenty-second to the fifty-seventh percentile. Separate ANOVAs on each of the six measures in Table 3 revealed significant differences on all except reading comprehension, where $p < .07$. Tukey HSD tests indicated that the percentile scores of the good students were significantly higher than those of the not-so-good students on all six achievement test measures, that the scores of the perfectionists were significantly higher than those of the not-so-good students on five of the six measures (all except reading comprehension), and that the performance of the perfectionists and good students did not differ significantly on any of the measures.

Another type of external validation was provided by the placement of children the
TABLE 3

ACHIEVEMENT TEST PERCENTILE SCORES OF PERFECTIONISTS, GOOD STUDENTS, AND NOT-SO-GOOD STUDENTS: EXPERIMENT 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Perfectionists</th>
<th>Good Students</th>
<th>Not-So-Good Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total math</td>
<td>86</td>
<td>81</td>
<td>37</td>
</tr>
<tr>
<td>Math computation</td>
<td>84</td>
<td>68</td>
<td>22</td>
</tr>
<tr>
<td>Math problem solving</td>
<td>80</td>
<td>80</td>
<td>38</td>
</tr>
<tr>
<td>Total reading</td>
<td>81</td>
<td>83</td>
<td>52</td>
</tr>
<tr>
<td>Word recognition</td>
<td>79</td>
<td>84</td>
<td>54</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>76</td>
<td>83</td>
<td>57</td>
</tr>
</tbody>
</table>

next year. Of the 36 children, three were retained in first grade and one was assigned to a learning disabilities class. Based on their performance on the experimental tasks, all four of these children had been classified as being in the not-so-good student group.

Interpreting the pattern of individual differences.—How can these differences among the perfectionists, good students, and not-so-good students be explained? Differences in two variables that influence performance within the strategy choice model—confidence criteria and peakedness of distributions of associations—provide a straightforward explanation. Specifically, the perfectionists appear to be children who set very high confidence criteria and have peaked distributions, the good students appear to be children who also have peaked distributions but who set less high confidence criteria, and the not-so-good students seem to be children who set lower confidence criteria and have less peaked distributions than children in the other two groups.

First consider the hypothesis that the perfectionists set higher confidence criteria than children in the other two groups. Within the strategy choice model, high accuracy and short solution times on retrieval trials will, other things equal, be accompanied by frequent use of retrieval. The performance of the perfectionists defied this pattern, though. As shown in Figure 2, they were the fastest and most accurate on retrieval trials, but also the least likely to use retrieval. The only way in which such a pattern would be produced by the strategy choice model is if the perfectionists set higher confidence criteria than children in the other two groups. This would lead to relatively infrequent use of retrieval because the only retrieved answers that would be stated would be ones with very high associative strength. Such answers, when they exist, would be likely both to be correct and to be retrieved quickly.

The perfectionists also appeared to have relatively peaked distributions of associations. The most direct reflections of peakedness are speed and accuracy on retrieval trials. The perfectionists were much faster and more accurate on retrieval trials than the not-so-good students, and somewhat more than the good students. Some of their superiority to the good students was due to their using retrieval on easier problems; however, this certainly does not explain their superiority to the not-so-good students. The safest conclusion seemed to be that the peakedness of the perfectionists’ distributions of associations was at least as great as that of the good students’ distributions, and that it was considerably greater than that of the not-so-good students’ distributions.

The not-so-good students appeared to have less peaked distributions than either of the other two groups, and also to set lower confidence criteria. First consider evidence that their distributions were less peaked. On all three tasks, the not-so-good students’ performance on retrieval trials was both slower and less accurate than that of children in the other two groups. Differences in problem difficulty accounted for only a small part of the differences between their performance and that of perfectionists, and almost none of the differences between their performance and that of the good students.

As discussed above, confidence criteria of the not-so-good students appeared to be much lower than those of the perfectionists; the not-so-good students retrieved on a higher percentage of trials, despite having less peaked distributions. At first glance, all of the differences between the not-so-good students and the good students might seem attributable to differences in the peakedness of their distributions. As would be expected from the view that the distributions of the not-so-good students were less peaked, they used retrieval on a lower percentage of trials and used it
more slowly and less accurately. However, the magnitudes of the differences suggested that they also probably set lower confidence criteria. As shown in Figure 2, the good students used retrieval on 10% more trials than the not-so-good students. However, they were correct on 28% more of retrieval trials, and their solution times on retrieval trials were less than half as long. The combination of less peaked distributions and lower confidence criteria would produce these relations.

In sum, this analysis suggests that perfectionists set higher confidence criteria than good students, who in turn set higher confidence criteria than not-so-good students. It also suggests that both perfectionists and good students have more peaked distributions than not-so-good students.

Two factors limited the assuredness with which these conclusions could be drawn, however. First, although the dimensions that differentiated individual performance were suggested by the strategy choice model, the model did not predict the particular individual differences that emerged. Simply replicating the findings with a different sample of children and different specific problems therefore seemed desirable.

The second issue involved methods for assessing strategies. In Experiment 1, the distinction between use of backup strategies and use of retrieval was made solely on the basis of overt behavior. If children produced audible or visible behavior relevant to solving the problem between the time the problem was presented and the time when they stated the answer, they were classified as using a backup strategy; if not, they were classified as using retrieval. This classification procedure had been used effectively with preschoolers' addition and subtraction in earlier studies (Siegler, 1987a; Siegler & Shrager, 1984) and therefore seemed reasonable in the present context. After Experiment 1 was conducted, however, another study (Siegler, 1987c) was conducted in which first graders were asked immediately after each addition problem how they had solved the problem. Results of this study provided two reasons for obtaining verbal reports of strategy use immediately following each problem. First, when asked how they had solved particular problems, the first graders often said that they used backup strategies on trials where they had not produced any overt behavior. Second, chronometric data strongly supported the validity of the children's verbal reports. Solution time and error patterns differed sharply on trials where children reported using different strategies, and the pattern of times yielded when children said they used each strategy conformed to what would have been expected from the operations inherently involved in executing that strategy.

Thus, it seems likely that relying solely on overt behavior as an index of backup strategy use underestimated the frequency of backup strategies and overestimated the frequency of retrieval. Further, it seemed possible that apparent differences among the three groups in percent use of retrieval actually reflected different rates of use of covert rather than overt backup strategies. Fortunately, the above-described results of Siegler (1987c) suggested that immediately retrospective verbal reports, when used together with videotapes of children's ongoing performance, would lead to more accurate classifications of strategy use. Therefore, both types of data were obtained in Experiment 2 and used to assess children's strategies.

**Experiment 2**

**Method**

**Participants.**—The children were 34 first graders, mean CA = 79 months (SD = 4.5 months). The group included 14 boys and 20 girls. They attended the same middle-class public school as the children in Experiment 1. The experimenter was a 28-year-old female research assistant.

**Problems.**—The problems used in this experiment were of similar numerical sizes and letter lengths as those used in Experiment 1 but differed in the particular problems and words. The 21 addition problems had larger addend sizes ranging from 2 to 14, smaller addend sizes from 1 to 6, and sums from 3 to 18. The 21 subtraction items had minuends ranging from 3 to 18, subtrahends from 2 to 8, and differences from 1 to 14. The 50 reading words ranged from two letters ("is") to nine letters ("breakfast"). As in Experiment 1, each word was printed individually on a 4 × 6-inch index card. Also as in Experiment 1, roughly 70% of the reading words and arithmetic problems had been encountered by the children in their textbooks; the other 30% were also sampled from the text but appeared in lessons that the class had not yet studied. The Metropolitan Achievement Test was taken by students 5 months after the experimental session. The same sub-tests were examined as in Experiment 1.

**Procedure.**—The procedure was almost identical to that followed in the first experiment. The only difference involved the
TABLE 4

SPEED AND ACCURACY ON EACH TASK: EXPERIMENT 2

<table>
<thead>
<tr>
<th>Task</th>
<th>% Retrieval</th>
<th>% Correct</th>
<th>Median RT</th>
<th>% Correct, Retrieval Trials</th>
<th>% Correct, Backup Strategy Trials</th>
<th>Median RT, Retrieval Trials</th>
<th>Median RT, Backup Strategy Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>49</td>
<td>84</td>
<td>5.4</td>
<td>92</td>
<td>76</td>
<td>3.6</td>
<td>7.1</td>
</tr>
<tr>
<td>Subtraction</td>
<td>35</td>
<td>70</td>
<td>10.7</td>
<td>85</td>
<td>63</td>
<td>3.1</td>
<td>12.1</td>
</tr>
<tr>
<td>Reading</td>
<td>52</td>
<td>54</td>
<td>7.6</td>
<td>82</td>
<td>26</td>
<td>1.9</td>
<td>28.4</td>
</tr>
</tbody>
</table>

request that children verbally describe the strategy that they used immediately after answering the problems. Children were told after the first problem, "We're interested in knowing how children 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 one or more follow-up questions. For example, if the child simply said "I counted," the experimenter would ask, "What number did you begin counting on?"

RESULTS AND DISCUSSION

Reliability of scoring of strategies using verbal reports as well as overt behavior was high. The two raters agreed on 99% of trials on the addition task, 97% of trials on the subtraction task, and 100% of trials on the word identification task. Cases of disagreement were resolved through discussion by the two raters.

Performance on Each Task

Averaged across all trials on all three tasks, speed and accuracy in Experiment 2 were similar to speed and accuracy in Experiment 1. Mean percent correct was 69%, versus 73% in Experiment 1. Median solution time was 7.9 sec, versus 8.0 sec in Experiment 1. In general, performance on the two arithmetic tasks was faster and more accurate in Experiment 2; performance on the word identification task was slower and less accurate (Table 4).

Use of verbal reports as well as overt behavior for strategy classification in Experiment 2 appeared to allow accurate detection of numerous covertly executed backup strategies. One consequence of correctly classify-
TABLE 5
CORRELATIONS OF INDIVIDUAL CHILDREN’S PERFORMANCE IN ADDITION,
SUBTRACTION, AND READING: EXPERIMENT 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Addition-Subtraction</th>
<th>Addition-Reading</th>
<th>Subtraction-Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>% retrieval</td>
<td>.76</td>
<td>.62</td>
<td>.48</td>
</tr>
<tr>
<td>% correct</td>
<td>.60</td>
<td>.28</td>
<td>.57</td>
</tr>
<tr>
<td>Mean RT</td>
<td>.70</td>
<td>.15</td>
<td>.24</td>
</tr>
<tr>
<td>% correct on retrieval trials</td>
<td>.52</td>
<td>.26</td>
<td>.47</td>
</tr>
<tr>
<td>% correct on backup strategy trials</td>
<td>.43</td>
<td>.30</td>
<td>.68</td>
</tr>
<tr>
<td>Mean RT on retrieval trials</td>
<td>.91</td>
<td>.41</td>
<td>.36</td>
</tr>
<tr>
<td>Mean RT on backup strategy trials</td>
<td>.08</td>
<td>.44</td>
<td>.30</td>
</tr>
</tbody>
</table>

Note.—All correlations are Pearson product-moment r’s. With df = 34, r’s > .33 are significant for p < .05.

Good Students, Not-So-Good Students, and Perfectionists

As in Experiment 1, a cluster analysis was conducted on each child’s percent retrieval, percent correct retrieval, and percent correct backup strategy use on each of the three tasks. Again, the three-group solution was considerably superior to the two-group solution. Separate ANOVAs on each variable indicated that in the two-group solution, the groups differed significantly on five of the nine variables. In the three-group solution, the groups differed significantly on all nine variables.

Next, to obtain converging validity for the results of the cluster analysis, performance of children in the three groups was contrasted on six measures that were not used as input to the analysis. These measures involved solution times on retrieval trials and solution times on backup strategy trials on each of the three tasks. Performance of the three groups differed significantly on four of the six measures: solution times on backup strategy trials on all three tasks, and solution times on retrieval trials in subtraction. Performance on the other two measures—solution times on retrieval trials in addition and reading—showed a similar pattern, and the differences among groups approached significance (p’s < .06 and .07, respectively).

The three groups yielded by the cluster analysis were sufficiently similar to those yielded by the analysis of the Experiment 1 data that it seemed appropriate again to call them the perfectionists, good students, and not-so-good students (Fig. 3). There were 14 children in the perfectionist group, 12 in the good student group, and eight in the not-so-good student group.

Differences between good students and not-so-good students closely followed those in Experiment 1. Tukey HSD tests indicated that the good students were correct significantly more often on both retrieval trials and backup strategy trials on all three tasks, that they were faster on backup strategy trials on all three tasks, that they were faster on retrieval trials on both the addition and subtraction tasks, and that they retrieved more often on both the addition and word identification tasks. All 15 of the differences were in the expected direction, and 13 of the 15 differences were significant.

The relation between perfectionists’ performance and that of children in the other two groups also resembled that of children in Experiment 1. Relative to the not-so-good students, the perfectionists were significantly more accurate on retrieval trials on all three tasks and on backup strategy trials on the addition and subtraction tasks. They also were significantly faster than the not-so-good students on backup strategy trials on all three tasks and on retrieval trials on the subtraction task. Despite this greater speed and accuracy, the perfectionists used retrieval significantly less often on the addition and subtraction tasks and nonsignificantly less often on the word identification task. Again, all 15 of the differences were in the expected direction; 11 of the 15 were significant.

The relation between performance of good students and perfectionists was similar though not identical to that in Experiment 1. As in Experiment 1, the speed and accuracy of the perfectionists and the good students was similar; only one of the 12 differences was significant (the good students were more accurate on backup strategy trials in word identification). Also as previously, the main difference between perfectionists and good students was that perfectionists used retrieval significantly less often than good students.
The differences were significant on all three tasks. However, unlike the case in Experiment 1, where directional differences in speed and accuracy of perfectionists and good students tended to favor the perfectionists, in Experiment 2 directional differences in speed and accuracy tended to favor the good students (Fig. 3).

Analyses like those in Experiment 1 were also conducted to determine whether differences among the three groups' speed and accuracy on retrieval trials could be explained by differences in the problems on which they used retrieval. The results indicated that the problems on which the perfectionists used retrieval were somewhat easier than the problems on which children in the other two groups did so. Averaged across the three tasks, the expected percent correct retrieval for the problems on which the perfectionists, good students, and not-so-good students used retrieval was 91%, 94%, and 85%, respectively. The three groups' observed percent correct retrieval was 92%, 90%, and 72%, respectively. Thus, the good students' percent correct retrieval exceeded by 6% the level expected from the problems on which they used retrieval, the perfectionists' percent correct retrieval exceeded by 1% that expected from the problems on which they used retrieval, and the not-so-good students' percent correct retrieval was 13% lower than would have been expected on the basis of the problems on which they used retrieval. Thus, after correcting for differences in problem difficulty, good students retrieved correctly on somewhat more trials (5%) than perfectionists, and both groups retrieved correctly on considerably more trials than not-so-good students (20% and 14%, respectively).

Sex differences.—As in Experiment 1, there were no significant sex differences among the three groups, \( \chi^2(2) < 1 \). Girls constituted 50% of the perfectionists, 67% of the
TABLE 6

<table>
<thead>
<tr>
<th>Measure</th>
<th>Perfectionists</th>
<th>Good Students</th>
<th>Not-So-Good Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total math</td>
<td>83</td>
<td>90</td>
<td>53</td>
</tr>
<tr>
<td>Math computation</td>
<td>78</td>
<td>88</td>
<td>54</td>
</tr>
<tr>
<td>Math problem solving</td>
<td>81</td>
<td>82</td>
<td>46</td>
</tr>
<tr>
<td>Total reading</td>
<td>61</td>
<td>88</td>
<td>45</td>
</tr>
<tr>
<td>Word recognition</td>
<td>62</td>
<td>81</td>
<td>43</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>60</td>
<td>86</td>
<td>48</td>
</tr>
</tbody>
</table>

good students, and 62% of the not-so-good students.

Achievement test performance.—As in Experiment 1, children's Metropolitan Achievement Test scores were used to provide a measure of the external validity of the classifications of children into three groups. The scores of 32 of the 34 children were used for this validation; one of the not-so-good students had missing subtests, and one of the perfectionists' scores included several subtests that could only be classified as outliers (e.g., a score more than 5 SD below the mean of the other children in the experiment).

The overall level of performance on the standardized tests was similar to that in Experiment 1. The mean total mathematics score was at the seventy-ninth percentile of the standardization sample, and the mean total reading score was at the sixty-eighth percentile (vs. the seventy-second and seventy-fourth percentiles in Experiment 1).

Also as in Experiment 1, differences between the achievement test scores of children in the three groups paralleled those in the experiment. ANOVAs indicated significant differences among the three groups on each of the six subtests (Table 6). The good students' average score on the six subtests ranged from the eighty-first to the ninetieth percentile, the perfectionists' average scores ranged from the sixtieth to the eighty-third percentile, and the not-so-good students' average scores ranged from the forty-third to the fifty-fourth percentile. The HSD comparisons indicated that the differences between good and not-so-good students were significant on all six measures. Differences between perfectionists and not-so-good students also were significant on the three mathematics achievement scales; on the three reading scales, the perfectionists scored higher, but none of the differences was significant. Finally, the good students scored significantly higher than the perfectionists on the reading comprehension and overall reading measures and scored nonsignificantly higher on the other four measures.

Interpreting the pattern of individual differences.—As in Experiment 1, differences among the groups labeled perfectionists, good students, and not-so-good students can be interpreted in terms of varying status on two underlying dimensions: stringency of confidence criteria and peakedness of distributions of associations. In particular, the children labeled perfectionists are viewed as having relatively peaked distributions and setting very high confidence criteria, the children labeled good students are also viewed as having relatively peaked distributions but setting less high confidence criteria, and the not-so-good students are viewed as children whose distributions are not very peaked and who do not set very high confidence criteria.

Differences in peakedness of distributions are especially evident in comparing the performance of the good and not-so-good students. The good students were superior on every measure that the model suggests is influenced by peakedness. They used retrieval more often on all three tasks, they were faster on retrieval trials on all three tasks, and they were more accurate on retrieval trials on all three tasks. The differences in speed and accuracy of performance on retrieval trials could not be attributed to differences in difficulty of problems where retrieval was used because the problems on which the good students used retrieval were equal in difficulty of retrieving correctly to those on which the not-so-good students did.

The perfectionists also appeared to have more peaked distributions than the not-so-good students. They performed more quickly and more accurately on retrieval trials on all three tasks. A portion of these differences was due to differences in problem difficulty; however, even after correcting for problem
difficulty, the perfectionists still retrieved considerably more accurately than the not-so-good students.

The good students' and perfectionists' high percentage of correct retrieval, 90% and 92%, respectively, suggested that both possessed peaked distributions on the problems they were presented. There were no significant differences between the two groups' speed and accuracy on retrieval trials. When problem difficulty was considered, directional differences tended to favor the good students. The safest conclusion, though, seems to be that children in both groups had relatively peaked distributions of associations, and that the degree of peakedness did not differ substantially.

Now consider differences in confidence criteria. The clearest contrast is in the performance of perfectionists and not-so-good students. As noted above, the perfectionists appeared to have more peaked distributions. However, the perfectionists used retrieval significantly less often. The pattern of differences between the two groups is difficult to explain without reference to some type of difference in threshold for stating retrieved answers.

The perfectionists also seemed to set higher confidence criteria than the good students. The good students appeared to possess slightly more peaked distributions of associations; when difficulty of problems on which they used retrieval was held constant, they were correct on 5% more trials. However, the difference in the two groups' percent use of retrieval was out of proportion to this small difference in peakedness. The good students used retrieval on 66% of trials, whereas the perfectionists used it on only 28%. None of the six differences between the two groups' speed and accuracy on retrieval trials was significant, but all three differences in percent use of retrieval were.

It was unclear whether the good students set higher confidence criteria than the not-so-good students. As noted already, the good students had more peaked distributions than the not-so-good students; they used retrieval more often, and were faster and more accurate than the not-so-good students on retrieval (and backup strategy) trials. Under these circumstances, detecting differences in confidence criteria requires a disproportion in the degree of difference on different measures. Whether such a disproportion existed was unclear. In terms of absolute percent retrieval and absolute percent correct retrieval, differences between the two groups were entirely proportional. The good students used retrieval on 20% more trials; they retrieved correctly on 19% more trials. There were no differences in problem difficulty to consider. However, solution times on retrieval trials of the two groups did seem disproportional; the mean solution time of the good students was 2.5 sec, versus 5.1 sec for the not-so-good students.

The main findings, though, were clear. As in Experiment 1, one group of children possessed peaked distributions and set very high confidence criteria, another group of children also possessed peaked distributions but set lower confidence criteria, and a third group of children possessed less peaked distributions than children in the other two groups and set low confidence criteria. Under these circumstances, it appeared reasonable to refer to the three groups by the same names as the groups in Experiment 1: perfectionists, good students, and not-so-good students.

General Discussion

The results of these experiments indicated that clear individual differences exist in first graders' strategy choices in addition, subtraction, and word identification. Differences were evident on two dimensions: knowledge of problems and thresholds for stating retrieved answers. Performance on standardized achievement tests lent external validity to the diagnoses of individual differences based on performance in the experimental situation. The basic depiction of individual differences was stable across two sets of problems, two groups of children, and two strategy assessment methods, one using verbal reports and one not doing so. Some of the issues raised by these findings are discussed below.

Generality.—The strategy choice examined in this study, the choice between statement of a retrieved answer and use of a backup strategy, is an extremely common one. It is not limited to arithmetic and word identification. Rather, it must be made on any task where people have enough experience with items to make retrieval of an answer a possibility and where they also possess one or more strategies to use if they cannot retrieve the answer. Some such choices, such as those involved in addition and subtraction, involve decisions between statement of a retrieved answer and use of an algorithmic procedure that always yields the correct answer if executed correctly. Three examples of such retrieval versus algorithmic backup strategy choices are deciding whether to tell time by
recognizing the clock setting or by counting from the hour, generating the number of days in June from memory or by reciting "30 days hath September," and deciding whether to write down a just-looked-up phone number or to simply keep it in memory. A special case of such algorithmic backup strategies involves reference books. By providing atlases, baseball record books, cookbooks, dictionaries, encyclopedias, farmers' almanacs, and myriad other sources, our culture enhances our opportunity to use algorithmic backup strategies when we otherwise would be at the mercy of our memories.

Other choices between retrieval and backup strategies involve nonalgorithmic backup strategies, where correct execution of the backup strategy does not guarantee success. Use of decoding on the word identification task in the present experiment is one such example; others include retrieving a spelling or relying on sound-symbol correspondences, trying to just remember a phone number versus rehearsing it, and remembering a recipe versus repeatedly tasting and seeing what the dish seems to need.

Thus, people often choose between stating retrieved answers and using backup strategies. The general similarity of the strategy choice does not guarantee that individual differences in choices on these varied tasks would parallel those in addition, subtraction, and word identification. However, it does seem likely that individual differences on these tasks could be analyzed in terms of the same dimensions as appeared to influence choices in the present study—knowledge of problems and stringency of confidence criteria.

Relations to other individual difference constructs.—The present study related individual differences in the experimental situation to individual differences in a standardized test situation. Both similarities and differences were apparent. One similarity was in the contrast between perfectionists and good students, on the one hand, and not-so-good students on the other. In both situations, performance of the perfectionists and good students was superior. The relation of the performance of the perfectionists to that of the good students, however, differed in the two situations. Although the two groups were comparable in achievement test performance, their strategies in the experimental situation differed substantially.

Both the similarities and the differences between analyses of standardized test and experimental performance can be interpreted straightforwardly. Individual differences detected in the experimental situation involved two dimensions: differences in knowledge of problems and differences in criteria for stating retrieved answers. Individual differences detected on the achievement tests involved only the first of these, differences in knowledge. This view explains why both experimental and standardized test performance would discriminate between not-so-good students and children in the other two groups, where the main difference is one of knowledge. Both analyses are sensitive to such differences in knowledge. It also explains why only the analysis of performance in the experimental situation would discriminate between perfectionists and good students, where the main difference is not one of knowledge but rather of confidence criteria, or more generally, cognitive style. As these differences between good students and perfectionists illustrate, there is more than one way to be smart.

The dimensions of individual differences that emerged in the present study seem intuitively to be related to another individual difference construct, reflectivity-impulsivity (Kagan, Rossman, Day, Albert, & Phillips, 1964). Kogan's (1983) definition of reflectivity-impulsivity as "the extent to which a child delays a response in the course of searching for the correct alternative" (p. 672) is clearly akin to the role of the confidence criterion in the decision whether to state a retrieved answer. Salkind and Wright's (1977) hypothesis that reflective and impulsive children differ on two dimensions, efficiency of information processing and impulsivity, seems especially similar to the present view of children as differing in knowledge of problems and confidence criteria.

There are also dissimilarities between the two analyses of individual differences. The most obvious is the existence of three rather than two groups in the present analysis. One likely reason for this dissimilarity is methodological. Classifications of reflectivity-impulsivity usually are based on median splits on speed and accuracy measures. Only children in two of the four cells created by the median splits (slow/accurate and fast/inaccurate) are typically examined, despite the increasing recognition that the other cells are also populated fairly heavily (Kogan, 1983). The present clustering analyses, in which no particular form was forced on the groupings of children, indicated that in these domains, at least, many children (the good students) have
high knowledge yet do not set especially high confidence criteria.

A second source of dissimilarities concerns the types of tasks that are studied. The tasks examined here are ones where children have considerable knowledge and experience. It seems likely that in any domain in which knowledge plays a large role, solution times and errors are likely to be positively correlated, as they were in the present study ($r$'s = .3 to .6 on most measures). In contrast, in the task most often used to measure refection-impulsivity, the Matching Familiar Figures Test, prior knowledge plays no direct role, and speed and accuracy correlate negatively, about $- .3$ to $- .6$ (Messer, 1976).

These differences notwithstanding, it seems desirable to study empirically the relation between the types of individual differences identified here and on measures of refection-impulsivity. Particularly intriguing is whether perfectionists and reflective students tend to be the same children, and whether not-so-good students and impulsives tend to be the same.

**Development of individual differences in strategy choices.**—Given that children differ in peakedness of distributions and confidence criteria, how might such differences develop? One likely contributor to differences in peakedness is accuracy of execution of backup strategies. Within the present strategy choice model, the more accurately that children execute backup strategies, the more peaked their distributions of associations become. Consistent with this assumption, on all three tasks in both experiments of this study, a child's percent correct in executing backup strategies correlated positively and significantly with the child's retrieval accuracy. In Experiment 1, the correlations were $r = .43$ for addition, $r = .55$ for subtraction, and $r = .57$ for reading. In Experiment 2, the correlations were $r = .42$ in addition, $r = .47$ in subtraction, and $r = .52$ in reading. In both experiments, perfectionists and good students were significantly more accurate than not-so-good students in executing backup strategies on addition and subtraction problems; the direction of the difference was the same in word identification. Beyond the context of the present experiments, individual differences in the backup strategies of sounding out words have been found to be an excellent predictor of individual differences in psychometrically assessed reading ability in the early grades (Curtis, 1980). Thus, ability to accurately execute backup strategies seems likely to contribute to the acquisition of peaked distributions of associations.

An educational implication of this view is that it might be useful to teach children, particularly not-so-good students, to more accurately execute backup strategies. This seems a relatively uncontroversial conclusion with regard to word identification, where decoding instruction is generally viewed as useful (Perfetti, 1985), especially for low-ability students (Chall, 1979). It seems more controversial in arithmetic, where many parents and teachers discourage children from using backup strategies such as counting fingers (Siegel & Shraga, 1984). However, the logic is similar. Teaching children to execute backup strategies more accurately affords them more opportunities to learn the correct answer (i.e., to build distributions with strong peaks at the correct answer). It also reduces the likelihood of associating incorrect answers, produced by faulty execution of backup strategies, with the problem. This logic, the positive correlations between accurate execution of backup strategies and retrieval accuracy, and the encouraging results of teaching accurate execution of word identification backup strategies, all argue for testing the effects of teaching children to more accurately execute backup strategies in arithmetic as well as in reading.

Less is known about the development of confidence criteria. One possibility is that setting of confidence criteria responds to previous experience using backup strategies. Use of backup strategies is more time-consuming than retrieval. If the backup strategies result in consistently successful performance, children may find it worthwhile to set high criteria, which will lead to their using backup strategies relatively often. If the backup strategies do not consistently yield correct performance, however, children may not find it worthwhile to take the greater time needed to use them. This interpretation is consistent with the finding that the not-so-good students execute backup strategies the least accurately and that they also set the lowest confidence criteria of children in the three groups. The view leads to the testable prediction that teaching children to more accurately execute backup strategies would also lead to their setting higher confidence criteria, resulting in more, as well as more accurate, use of backup strategies.

**Cognitive models and individual differences.**—The present research on individual differences was guided by a previously for-
mulated cognitive model. Within the model, the two main influences on choices of whether to state a retrieved answer to a problem are the peakedness of distributions of associations on different problems and the stringency of confidence criteria. Data from the present study indicate that variability along these two dimensions also contributes to individual differences, specifically, differences in first graders’ addition, subtraction, and word identification. The individual differences identified in the experimental situations were related to those evident on achievement tests but went beyond them to capture differences not apparent in the standardized test scores. The findings illustrate how detailed cognitive models can contribute to understanding of individual differences.

References


