Doing as they are told and telling it like it is: Self-reports in mental arithmetic

BRENDA L. SMITH-CHANT and JO-ANNE LEFEVRE
Carleton University, Ottawa, Ontario, Canada

Adults ($n = 64$) solved single-digit multiplication problems under both speed and accuracy instructions. Half also provided self-reports of their solutions to the problems. The participants with relatively low levels of arithmetic fluency were most influenced by instructional requirements. They responded more slowly and accurately when asked to provide descriptions of their solution procedures, whereas the performance of the participants with high and average levels of arithmetic fluency did not change. Furthermore, the performance of the low-fluency participants was more affected by speed and accuracy demands than was that of the other individuals, but only when the low-fluency participants were also required to provide self-reports. Accordingly, models of mental arithmetic will need to include roles for individual differences and situational factors.

Can adults consistently and accurately describe how they solve simple arithmetic problems, such as 3×4 ? Research based on verbal reports suggests that adults use multiple approaches to solve such problems (Geary, Frensch, & Wiley, 1993; Geary & Wiley, 1991; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996; Svenson, 1985). These findings present a challenge to existing theories in which direct retrieval from memory is assumed to be the sole solution procedure used by adults to solve simple arithmetic problems (e.g., Ashcraft, 1992; Campbell, 1995; cf. Baroody, 1984). Kirk and Ashcraft (2001) suggested that the self-report methodology used to study adults' solution procedures has three major shortcomings. First, people may change their behaviors when they are asked to describe what they are doing. Second, people may be unable to accurately report their solution procedures, and thus self-reports may not be valid descriptions of mental processing. Third, aspects of the experimental procedures may bias the kind and types of verbal reports and solution procedures that participants report. For example, providing participants with details about types of possible solution procedures that could be used to solve arithmetic problems or even mentioning that different approaches to such problems are possible could bias participants to report multiple solution procedures. In

the present research, we provide evidence that, although self-reports influence the behavior of some individuals, such reports provide valid and useful indices of solvers' mental procedures. Furthermore, we contend that information about the variability in selection of procedures that can be gained from self-reports may be crucial for developing comprehensive models of mental arithmetic.

In a variety of studies, educated adults reported using solution procedures other than direct retrieval from memory on simple addition, subtraction, multiplication, and division problems (Campbell & Xue, 2001; Geary et al., 1993; Geary & Wiley, 1991; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre & Morris, 1999; LeFevre, Sadesky, & Bisanz, 1996; LeFevre, Smith-Chant, Hiscock, Daley, & Morris, 2003). For example, adults reported that they sometimes solved problems such as $9 + 2$ by counting 10, 11, solved problems such as 9×6 by transforming the problem to $10 \times 6 - 6$, solved problems such as $15 - 9$ by solving $15 - 10 - 1$, or solved problems such as $56 \div 8$ by transforming it to $8 \times _ = 56$. Typically, the procedures used by adults to produce the solution involved multiple steps, including retrieval of a well-known fact and the use of arithmetic principles.

The frequency with which adults reported using procedures other than direct retrieval varied across studies, across individuals, and across problems. People reported using procedures other than direct retrieval more for addition and subtraction than for multiplication and more on problems with larger operands (e.g., 7×8) than on problems with smaller operands (e.g., 2×5 ; Campbell & Xue, 2001; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996; LeFevre et al., 2003). Individuals who responded more quickly and accurately reported using retrieval more frequently than did individuals who responded more slowly and less accurately (Hecht, 1999; LeFevre, Sadesky, & Bisanz, 1996). Indi-

This research was supported by the Natural Sciences and Engineering Research Council of Canada through a graduate scholarship to B.L.S.-C. and through a research grant to J.-A.L. We thank Chris Herdman, Ben Coleman, Dawn Mullins, and Diana DeStefano for their helpful comments on earlier versions of this article. We also acknowledge J. I. D. Campbell, M. Ashcraft, and J. Zbrodoff for their comments on later versions of this article. Correspondence concerning this article should be addressed to B. L. Smith-Chant, Department of Psychology, Trent University, Peterborough, ON, K9J 7B8 Canada (e-mail: bresmith@trentu.ca or jo-anne_lefevre@carleton.ca).

viduals educated in China reported using retrieval more frequently than did individuals educated in North America (Campbell & Xue, 2001; LeFevre & Liu, 1997).

The use of self-report information for studying mental arithmetic in adults follows on the strong tradition of observing the solution procedures of children (Lemaire & Siegler, 1995; Siegler, 1987, 1988a, 1988b; Siegler & Jenkins, 1989; Siegler & Shipley, 1995; Siegler & Shrager, 1984). However, despite general agreement that use of multiple procedures bears on the theoretical understanding of adults' arithmetic cognition (Ashcraft, 1995), the issue of whether self-reports are useful sources of information about mental arithmetic has been studied directly by only a few researchers (Cooney & Ladd, 1992; Kirk & Ashcraft, 2001; LeFevre et al., 2003; Russo, Johnson, & Stephens, 1989).

To address the issue of whether participants are reactive to self-report demands—that is, whether they change their behavior when they are required to provide self-reports—researchers have compared the performance of individuals who provided self-reports with that of individuals in silent control conditions. Generally, there are few differences in patterns of latencies or errors between self-report and silent control groups, although those conclusions may depend on such factors as the form of the instructions or the particular arithmetic operation that was tested (Cooney & Ladd, 1992; Kirk & Ashcraft, 2001; LeFevre & Morris, 1999; LeFevre et al., 2003). Kirk and Ashcraft used a different approach to examine whether self-report demands influence behavior. In addition to a silent control condition and a self-report condition, they included two biasing conditions in which participants were encouraged either to use memory retrieval (the retrieval bias condition) or to use procedures (the procedure bias condition). If participants are reactive to the requirement to report their solution procedures, their performances should be associated with the type and degree of biasing they experience.

In Kirk and Ashcraft's (2001) first experiment, participants solved single-digit addition problems in one of the four instructional conditions. Biasing instructions had a significant influence on the participants' self-reports, in that the participants in the retrieval bias condition reported using retrieval on 90% of the trials (on average), whereas the participants in the procedure bias condition reported retrieval on only 32% of the trials. The standard self-report group reported using retrieval on 50% of the trials. In the second experiment, biasing instructions had a similar effect on self-reports for multiplication solutions, so that retrieval was reported on 96%, 62%, and 71% of the trials in the retrieval bias, procedure bias, and standard self-report conditions.

Kirk and Ashcraft's (2001) primary goal was to explore the consequences of deliberate biasing. Their results suggest that participants are reactive, in that they can be biased toward either using or reporting procedures in accord with the instructional demands of the experimental condition. It is less clear, however, whether

the self-reports were veridical. For addition, self-reports were generally consistent with the participants' behavior, as indexed by patterns of latencies. For multiplication, however, self-reports appeared to be unrelated to overall latencies when these were averaged across reported procedures. When latencies on individual trials were examined, however, they corresponded closely to the self-reports on those trials. Furthermore, the procedure reports that increased the most from the retrieval bias condition to the procedure bias condition, counting (from 2% to 19%), had latencies that were as fast as or faster than those on retrieval trials. Thus, averaging latencies across procedures and comparing these latencies across conditions is unlikely to provide much information about whether the self-reports were veridical reflections of behavior.

Other aspects of Kirk and Ashcraft's (2001) results suggest that the biasing instructions may have produced patterns of behavior on multiplication problems that were quite different from that observed in other experiments for which self-reports were collected. In particular, the frequency with which the participants reported retrieval in Experiment 2 (multiplication) increased as the size of the operands increased. In other research, participants have reported using retrieval less (and non-retrieval procedures more) as problem size increased (Campbell & Xue, 2001; Hecht, 1999; Kirk & Ashcraft, 2001, Experiment 1; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). This puzzling pattern of results suggests that Kirk and Ashcraft's conclusions about the effects of biasing may be limited to certain types of experimental conditions. Nevertheless, it seems clear that the form and content of instructions are critical in studies in which introspective techniques are used (Cooney & Ladd, 1992; Russo et al., 1989).

Kirk and Ashcraft (2001) also observed substantial individual differences in the percentage of trials that the participants solved by direct retrieval. In the retrieval bias group, the percentage of multiplication retrieval trials ranged from 77% to 100%, and in the procedure bias group, the percentage of multiplication trials in which retrieval was reported ranged from 0% to 100%. Clearly, there were substantial differences in the susceptibility to bias across individuals. In accord with the view that participants vary in their responsiveness to task or situational demands, we hypothesized that the effects of instructional emphases will depend greatly on the characteristics of the individuals who participate in the research. Knowledge of how variations in arithmetic fluency may influence participants' solution approaches may be crucial for understanding the influence of self-reports on performance (e.g., Campbell & Xue, 2001; Hecht, 1999; LeFevre & Kulak, 1994; LeFevre, Sadesky, & Bisanz, 1996).

The issue of whether participants' behavior is reactive to self-report requirements raises more general questions about the impact of other instructional manipulations on participants' performance. In arithmetic studies, re-

searchers vary on whether they emphasize speed or accuracy in the instructions to participants. When speed is emphasized (e.g., Campbell, 1994, 1995), participants tend to respond more quickly and make more errors than in similar studies in which accuracy is emphasized (e.g., LeFevre & Liu, 1997; LeFevre & Morris, 1999). Campbell and colleagues (Campbell, 1995; Campbell & Graham, 1985; Campbell & Oliphant, 1992) have suggested that requiring participants to respond quickly discourages them from using nonretrieval solution procedures and increases the probability that they will answer arithmetic problems by retrieving the solution from memory. However, researchers have not directly compared the impact of differentially emphasizing speed and accuracy requirements on the solution procedures reported by adults.

In the present research, participants solved single-digit multiplication problems. We manipulated speed and accuracy instructions in conjunction with self-report and silent control conditions. The participants were divided into skill groups based on an independent measure of arithmetic fluency. High-skill participants have been assumed to use direct retrieval consistently, quickly, and accurately across a wide range of problems and task conditions. Low-skill participants have been assumed to use retrieval less efficiently and to select from among a wider range of potential solution procedures. Thus, we predicted that, in contrast to high-skill participants, less-skilled participants would (1) use retrieval less frequently, (2) be more reactive to self-report requirements because they are more likely to use and report using nonretrieval solution procedures, and (3) show greater range of latencies for speed versus accuracy instructions as they switched from slower solution procedures to faster procedures for some problems. We also tested the hypothesis that participants would report using retrieval more frequently under speed instructions than under accuracy instructions.

METHOD

Participants

Sixty-four students (32 males and 32 females) were recruited from introductory psychology classes. The participants received either credit toward partial fulfillment of a course requirement or a \$12 honorarium. The participants ranged in age from 19 to 41 years, with a median age of 21.

Materials

Multiplication production task. The set of 64 multiplication problems used in this task included all possible combinations of single-digit multiplication problems from 2×2 to 9×9 . The 0- and 1-operand problems were excluded from the production task because there is considerable empirical evidence suggesting that these problems are solved using fast and efficient rules (Ashcraft, 1992). The participants answered each problem four times, twice under instructions in which speeded response was emphasized and twice under accuracy instructions.

Two separate stimulus lists were created with different orders of problems. Within each of the lists, the order of items was semi-randomized, with each presentation of a problem occurring only once in each half of the list.¹ Furthermore, problems with identical

answers (e.g., 3×4 , 6×2 , 4×3 , and 2×6) were separated by a number of intervening trials, and large (products greater than 26) and small (products less than 26) problems were interspersed equally throughout each list. The participants solved one of the lists of problems with instructions emphasizing speed and the other list with instructions emphasizing accuracy. The order of presentation of the speed and accuracy conditions and list order were counter-balanced across participants.

Arithmetic fluency task. The participants completed the addition and the subtraction–multiplication subtests of the French Kit (French, Ekstrom, & Price, 1963). Each subtest of this pencil-and-paper task consists of two pages of problems. During the task, the participants were required to solve as many problems on a page as possible in 2 min. The addition subtest consisted of three two-digit numbers arranged vertically in a column. The subtraction and multiplication subtest required the participants to complete a row of two-digit by two-digit subtraction problems (e.g., $48 - 19$) alternating with a row of two-digit by one-digit multiplication problems (e.g., 14×3). The total number of problems correctly solved across the four pages provides a measure of arithmetic fluency (i.e., the ability to solve arithmetic problems quickly and accurately).

Mathematics background and interests questionnaire. The participants answered questions concerning their age, sex, educational background, and previous educational training in mathematics. They were asked to indicate the extent to which they use non-retrieval procedures to solve basic multiplication, addition, and subtraction problems, such as derived facts, counting, rules, and tricks (e.g., songs or rhymes). The participants also completed portions of the Mathematics Skills Questionnaire used by LeFevre, Kulak, and Heymans (1992). A comprehensive analysis of how the questionnaire data is related to participants' performance across a series of studies is reported in LeFevre et al. (2003). Hence, questionnaire responses are not discussed further in the present paper.

Procedure

The participants were tested individually in a single session lasting approximately 1 h. Each participant was seated comfortably in front of an amber monochrome computer monitor and was fitted with a headset containing a microphone with a voice-activated timing switch. In all conditions, verbal latencies were recorded to the nearest millisecond, and the experiment was controlled using an 80286 IBM-type computer. The experimenter was always present to enter the participant's responses and indicate whether the voice key had triggered properly. The participants were asked to read a set of instructions provided on the computer screen. In the speeded condition, the participants were given instructions similar to those described by Campbell (1995):

You are being tested on how quickly you can solve simple multiplication problems. First, you will see an asterisk in the centre of the computer screen. The asterisk will begin to flash. This signals that the asterisk will be replaced by a single-digit multiplication problem like ' 3×4 '. I would like you to say the answer as quickly as you possibly can, without making any mistakes. Occasional mistakes are normal when people go fast; so do not be too concerned if you make a mistake. It is important that you respond as quickly as possible.

In the accuracy emphasis condition, instructions comparable to those described by LeFevre, Bisanz, et al. (1996) were used:

You are being tested on how accurately you can solve multiplication problems. First, you will see an asterisk in the centre of the computer screen. The asterisk will begin to flash. This signals that the asterisk will be replaced by a single-digit multiplication problem, like ' 3×4 '. I would like you to say the correct answer as quickly as you can. Occasional errors are normal, but please try to avoid making mistakes. It is important that you respond as accurately as possible.

Half of the participants ($n = 32$) completed the two tasks with the additional requirement to describe the solution procedure they used

to solve each problem. The additional instructions in this condition were similar to those used by LeFevre, Bisanz, et al. (1996), with the removal of the descriptions of particular procedures:

After you say the answer, the words 'How did you solve the problem?' will appear on the screen. I would like you to explain how you arrived at your answer. For example, you might 'just know' or 'just remember' the answer. If so, tell me you remembered or retrieved the answer. Remembering is just one way to solve a multiplication problem. Sometimes people figure out the answer by changing, or simplifying the problem. Please tell me exactly how you solved each problem. If you used more than one solution approach, please tell me everything that you did.

Once an answer response was made, the cue, "How did you solve the problem?" appeared at the fixation point. The participants were encouraged to report as much as possible about how they solved each problem, and the experimenter recorded their descriptions. Note that, although the instructions used in the present research reduced the potential for demand-induced biases in procedure reports that might be caused by detailing specific types of solution procedures, the instructions still conveyed a level of expectation that the participant would be likely to report solution procedures other than retrieval. Accordingly, these instructions cannot be considered as completely free from potential demand bias (cf. Kirk & Ashcraft, 2001).

After the participants read the instructions, they completed 10 practice trials. On each trial, an asterisk was presented for 1 sec to indicate the fixation point for the stimulus. This asterisk then was removed from the screen for 250 msec, was presented for 250 msec, was removed for 250 msec, was presented for 250 msec, and was removed for 250 msec. This gave the asterisk the appearance of flashing twice. The problem was then presented on what would have been the third flash, with the operation sign appearing at the fixation point. The problem remained on the screen until the participant responded or until 15 sec had elapsed. Upon detection of a verbal response (or after 15 sec), the problem disappeared, and the computer recorded the response latency. The experimenter then recorded the response (e.g., the stated answer, a failure of the voice-activated relay, or no response). Once the experimenter pressed the Enter key, the next trial was initiated. The average intertrial interval was 2.6 sec for the silent control condition and 4.6 sec for the self-report condition. The longer intertrial interval for the self-report condition reflects the additional time required by the participant to report a solution strategy and the time required for the experimenter to record that solution procedure. No feedback about accuracy or response time was provided to the participants. The French Kit was administered between the speed and the accuracy conditions to provide the participants with a rest from the computerized task. The questionnaires were given last, to avoid biasing the participants' self-reports during the production task.

Coding of Solution Procedures

During the experiment, solution procedures reported on each trial were coded into three mutually exclusive categories: direct retrieval, invalid response, or other. A solution procedure was classified as direct retrieval if the participant said something like "I just knew that" or "I have memorized the answer to that problem" and they did not report using another solution procedure. If the participant made a nonanswer vocalization (e.g., *um*) or the response failed to be registered by the computer, the response was coded as invalid, and the reason for the invalidation was recorded. Any other response was classified as *other*, and details were recorded to allow a subsequent detailed classification of solution reports.

RESULTS

Skill Groups

Total number correct on the four pages of the French Kit fluency test was calculated for each individual. Typically, samples of undergraduates have mean scores of approximately 80 ($\sigma = 20$) on this test (LeFevre et al., 2003). In this experiment, mean fluency score was 81 ($SD = 23$). The participants were categorized as low, average, or high skill, according to their fluency score. As is shown in Table 1, low-skill participants had fluency scores less than 70 (i.e., one half of a population standard deviation below the expected mean), average-skill participants had scores between 70 and 90, and high-skill participants had scores of 90 or above (i.e., one half of a population standard deviation above the expected mean). The fluency scores were analyzed in a 2 (report group: silent control or self-report) $\times$ 3 (skill: low, average, or high) analysis of variance (ANOVA). The only significant effect was for skill group [$F(2,58) = 121.53, MS_e = 100.52, p < .01$]. Skill group was used as an index of individual differences in all of the following analyses.

Analyses of Latencies and Percentages of Errors

The participants solved a total of 8,192 problems. Of these, 518 were errors, and 343 were invalid. The percentage of invalid trials was similar in the speed and the accuracy conditions (4.7% and 3.7%), whereas the participants made more errors in the speed condition than in the accuracy condition (8.6% vs. 4.0%). Invalid trials were not analyzed.

Table 1
Performance on the Multidigit Arithmetic Test (Number Correct)
by Report Condition (Self-Report, Silent Control) and Skill Group

Measure	Skill Group					
	High		Average		Low	
	Silent Control	Self-Report	Silent Control	Self-Report	Silent Control	Self-Report
<i>N</i>	13 (7)*	7 (1)	9 (3)	13 (7)	10 (6)	12 (8)
Mean	112	106	78	79	60	58
<i>SD</i>	15	13	5	6	5	10
Minimum	90	91	71	71	53	39
Maximum	137	128	88	88	67	68

*Numbers in parentheses indicate the number of females in each group.

Mean latencies on correct trials and mean percentages of errors on valid trials were analyzed in separate 3 (skill: low, average, or high) $\times$ 2 (report group: silent control or self-report) $\times$ 2 (instructional bias: speed or accuracy) $\times$ 4 (problem size: very small, small, large, or very large) ANOVAs, with repeated measures on the last two factors.² Problem size categories were in accord with those used by Kirk and Ashcraft (2001): Very small problems had products of 15 or less ($n = 16$), small problems had products between 16 and 25 ($n = 16$), large problems had products between 27 and 42 ($n = 17$), and very large problems had products of 45 or greater ($n = 15$). F and MS_e values for these analyses are shown in Table 2. Because latencies and percentages of errors showed complementary patterns, these dependent results are discussed together. The results discussed were significant at $p < .05$, unless otherwise indicated. The 95% confidence intervals shown in the figures were calculated using the approach recommended by Loftus and Masson (1994). The data for latencies and errors across all conditions are shown in Figure 1.

Consistent with all other research on simple arithmetic, latencies and percentages of errors (in parentheses) increased with problem size: 907 msec (1%), 1,095 msec (3%), 1,367 msec (10%), and 1,567 msec (18%) for very small, small, large, and very large problems, respectively. Latencies and errors varied with skill group in such a way that latencies and errors increased as skill decreased. Mean latencies (percentages of error) were 936 msec (5%), 1,113 msec (7%), and 1,653 msec (13%) for high-, average-, and low-skill participants, respectively. These

findings are consistent with other research in which performance on single-digit arithmetic problems was correlated with performance on arithmetic tasks that required multidigit arithmetic (Campbell & Xue, 2001; Hecht, 1999; Kirk & Ashcraft, 2001; LeFevre, Bisanz, et al., 1996). Skill also interacted with problem size for both latencies and percentages of errors, as is shown in Figure 1. The slope of the increase in latencies and errors with problem size was much greater for the low- than for the average- or high-skill participants.

The participants responded more quickly when given speed instructions than when given accuracy instructions (1,126 vs. 1,342 msec) and made more errors (10% vs. 6%), indicating that they were responsive to the instructional biases and showed a speed-accuracy tradeoff. Instructional bias also interacted with problem size for both latencies and errors, as is shown in Figure 1. The increase in latencies with problem size was larger in the accuracy bias condition than in the speed bias condition. Thus, latencies were reactive to instructional bias in such a way that the participants responded relatively more quickly on larger problems under speed instructions than under accuracy instructions. Importantly, the attenuation in the problem size effect in the speed condition was accompanied by an increase in errors on the large and very large problems. Instructional bias, therefore, had differential effects across problem size.

The participants in the silent control condition made more errors than did those in the self-report condition (10% vs. 7%), and there was a corresponding trend in latencies, so that the participants in the silent control con-

Table 2
Analyses of Variance for Analysis of Solution Latency and Percentage of Error: 3 (Skill: High, Average, or Low) $\times$ 2 (Report Group: Silent Control or Self-Report) $\times$ 2 (Instructional Bias: Speed or Accuracy) $\times$ 4 (Problem Size: Very Small, Small, Large, or Very Large)

Effects	df	F Values	
		Latency	% Error
Between Subjects			
Report group	1	3.07*	9.42***
Skill	2	17.42***	20.69***
Report group $\times$ skill	2	1.64	0.38
MS_e	58	(1,319,825)	(140.0)
Within Subjects			
Problem size (PS)	3	56.64***	125.08***
PS $\times$ skill	6	6.02***	11.74***
PS $\times$ report group	3	6.76***	4.70***
PS $\times$ report group $\times$ skill	6	3.26***	1.42
MS_e	174	(182,880)	(58.9)
Instructional bias	1	43.39***	37.65***
Instructional bias $\times$ skill	2	3.03*	0.48
Instructional bias $\times$ report group	1	0.37	0.65
Instructional bias $\times$ skill $\times$ report group	2	0.38	2.97*
MS_e	58	(130,788)	(52.3)
PS $\times$ instructional bias	3	17.24***	7.52***
PS $\times$ instructional bias $\times$ skill	6	1.27	0.91
PS $\times$ instructional bias $\times$ report group	3	1.77	0.84
PS $\times$ instructional bias $\times$ skill $\times$ report group	6	1.76	1.43
MS_e	174	(33,494)	(27.6)

* $p < .09$. *** $p < .01$

Figure 1. Latencies (line graphs) and percentages of errors (bar graphs) across skill (high, average, or low), instructional bias (speed or accuracy), problem size (very small, small, large, or very large), and report groups (self-report or silent control). For latencies, 95% confidence intervals were calculated on the basis of separate analyses for each skill group because the three-way interaction shown is significant and within-group differences are of primary interest. For errors, 95% confidence intervals were calculated on the basis of the MS_e for the four-way interaction that is shown.

dition responded more quickly than those in the self-report condition (1,143 vs. 1,325 msec; $p = .085$). The observation that participants in self-report conditions tend to make fewer errors and take more time to solve problems than participants in silent control conditions has been noted in previous studies (Cooney & Ladd, 1992; LeFevre et al., 2003). LeFevre et al. (2003) suggested that the tendency for participants to respond more carefully in the self-report condition might occur because they want to avoid having to explain how they arrived at an incorrect answer.

The impact of self-reports on performance was further modified by problem size. For latencies, this interaction is subsumed by the significant three-way interaction of

problem size, report group, and skill (see Figure 1). For completeness, Figure 1 presents the three-way interactions for both speed and accuracy instructions, but the patterns are similar across instructional conditions. The latencies of the high-skill participants were not different across report groups. For errors, the high-skill participants in the silent control group made more errors than did those in the self-report group only on large problems under speed biasing instructions. Hence, as was predicted, the high-skill participants were minimally reactive to the requirement to provide self-reports.

The average-skill participants were moderately reactive to the self-report requirements. The self-report participants tended to respond more slowly than the silent control

participants, but the difference was significant only for large problems under speed bias instructions. The average-skill participants in the silent control group made more errors than did the participants in the self-report group on both large and very large problems in both instructional conditions. Thus, the average-skill participants generally maintained their response latencies across report conditions but showed a decrease in errors on larger problems when they were instructed to provide self-reports.

In contrast to the two other groups, the low-skill participants were very reactive to the self-report requirements. As is shown in Figure 1, the participants in the self-report condition showed much greater increases in latencies and errors with problem size than did the participants in the silent control condition. Although the increase was even larger under speed than under accuracy bias instructions, the pattern of significant differences was the same. Thus, for the silent control group, the problem size effect in latencies was greatly reduced for large and very large problems. Similarly, patterns of errors were reactive for these participants. They showed a decrease in errors under self-report requirements on small, large, and very large problems in the accuracy bias condition and on very large problems in the speed bias condition. Thus, these low-skill participants showed very clear tradeoffs between accuracy and speed when instructional bias and report requirements were varied. As will be detailed below in the section on self-reports, the differences across speed and accuracy bias instructions for the low-skill participants were systematically related to their selection of procedures in these conditions.

Summary. As was predicted, the participants responded to self-report requirements in ways that varied with arithmetic fluency. The high-skill participants showed minimal reactivity when asked to provide self-reports. They also showed smaller effects of speed versus accuracy instructions than did the average- or the low-skill participants. In contrast, the low-skill participants were most reactive to task requirements. They showed larger differences between speed and accuracy conditions than did the high- or the average-skill participants. Most important, when they were asked to provide self-reports, the low-skill participants responded much more slowly on large and very large problems and made fewer errors than did the similarly low-skilled individuals who did not give self-reports. Thus, the results of this experiment support the hypothesis that participants' responses to instructional and task requirements will depend on their arithmetic skill.

Analysis of Self-Reports

The participants' reports of their procedures (other than retrieval) were classified into four categories: (1) derived facts, (2) counting or addition, (3) miscellaneous, and (4) guessing (LeFevre, Bisanz, et al., 1996). Retrieval included all of the trials on which the participants reported that they retrieved the answer from memory or "just knew it." Derived facts included all procedures that

involved use of a known fact to calculate the presented answers, as in solving 9×6 as $10 \times 6 - 6$. Counting or addition included trials on which the participants added, as in solving 2×6 as $6 + 6$ or 3×5 as $5 + 5 + 5$ or trials on which they reported counting up by increments of one of the operands, as in solving 3×5 as 5, 10, 15. *Miscellaneous* included all other procedures, such as using a *nines* rule or trick (e.g., 9×6 is 54, because $6 - 1$ is 5, and $9 - 5$ is 4) and idiosyncratic procedures, such as songs. Guessing was recorded when a participant said, "I guessed."

The decision to classify *guessing* as a procedure was based on a number of considerations. When the participants were asked whether a guess was the same as *just remembered*, some participants indicated that they considered an answer obtained by *guessing* as distinct from one obtained directly from memory. As well, a participant's indication that an answer was solved via guessing was not simply a justification for an incorrect response. The participants more often stated that they "just remembered" than that they "guessed" when answers were incorrect and guesses were not necessarily incorrect, as will be shown below. Thus, it was not clear whether guessed answers reflected responses generated using direct retrieval or some other solution process. Accordingly, we decided not to arbitrarily reclassify the participant's spontaneously generated solution reports of guessing as direct retrieval. The distribution of procedure reports in the speed and accuracy conditions is shown in Table 3. Retrieval was used most frequently, followed by derived facts, guessing, counting/addition, and miscellaneous procedures.

The percentage use of each of the retrieval, guessing, derived facts, and counting/addition procedures was calculated for each participant in the speed and accuracy bias conditions by the four categories of problem size. For each category of procedure, mean percentage of reported use was analyzed in a 3 (skill: high, average, or low) $\times 2$ (instructional bias: speed or accuracy) $\times 4$ (problem size: very small, small, large, or very large) ANOVA, with repeated measures on the last two factors. Note that these analyses are not independent, in that an increase in the percentage of use of one solution procedure is necessarily associated with a decreased use of other solution procedures. F and MS_e values for these analyses are shown in Table 4, and means are shown in Figure 2 for retrieval and in Figure 3 for counting/addition, derived facts, and guessing. Miscellaneous procedures were not analyzed, since they represented a very diverse set of solutions.

Retrieval. The participants' reports of retrieval did not vary with instructional emphasis ($F < 1$). Thus, the participants did not alter the frequency with which they reported using retrieval as a function of the speed and accuracy instructions (see Figure 2). Instead, instructional biases influenced the participants' selection of non-retrieval procedures, as will be described below. As was expected, percentage of reported use of retrieval decreased with problem size: 96%, 90%, 83%, and 75%.

Table 3
Prevalence, Latencies, and Accuracy of Procedures

Reported Procedure	Trials (%)	<i>n</i> *	Frequency of Use (%)†	Latency‡		Percentage of Errors
				<i>M</i>	<i>SD</i>	
Accuracy						
Retrieval	86.0	32	54–100	1,169	702	2.7
Derived facts	6.9	18	1–32	4,017	2,640	5.0
Guessed	2.2	12	1–20	4,021	2,838	16.0
Counting/addition	2.8	11	1–19	2,697	2,779	2.7
Miscellaneous	2.2	7	3–18	1,870	850	0.1
Speed						
Retrieval	86.0	32	46–100	1,023	568	5.3
Derived facts	4.3	17	1–27	3,644	2,618	6.8
Guessed	6.0	17	1–37	2,886	1,982	33.6
Counting/addition	2.6	12	1–15	1,182	956	2.3
Miscellaneous	1.2	7	1–13	1,423	729	1.6

*Number of participants who reported using the procedure at least once. †Range across participants who used the procedure at least once. ‡Correct trials only (in milliseconds).

This finding is consistent with other results in the literature (Hecht, 1999; LeFevre, Bisanz, et al., 1996).

As was predicted, the low-skill participants reported using retrieval less frequently (77%) than did the average- (90%) or the high-skill (91%) participants. Reported use of retrieval varied with problem size and skill, as is shown in Figure 2. The pattern of reported retrieval use was very similar for the high- and the average-skill participants across problem size. The low-skill participants did not differ from the other two groups on the very small problems but reported less retrieval on small, large, and very large problems. This pattern of procedure selection is consistent with the view that individuals who are skilled at simple arithmetic use memory retrieval to solve arithmetic problems (Campbell & Xue, 2001; Hecht, 1999; LeFevre & Liu, 1997; LeFevre, Sadesky, & Bisanz, 1996). In summary, the participants did not change the frequency with which they reported using di-

rect retrieval as a function of task demands. Instead, reports of retrieval varied with the participants' arithmetic fluency.

Nonretrieval procedures. In support of the view that self-reports are veridical reflections of mental processes, the participants' reports of procedures other than retrieval varied with problem size. As is shown in Figure 3, the participants reported more use of guessing and derived facts as problem size increased. In contrast, the participants reported less use of counting and addition procedures with increases in problem size, presumably because counting and addition procedures are cumbersome and unreliable when a large number of increments are required.

There were no two-way interactions between problem size and skill. However, patterns of reported procedure use varied with instructional bias. The participants reported using more guessing with speed than with accu-

Table 4
Analysis of Variance Summary Information for Analyses of Percentage Reported Use of Procedures: 3 (Skill: High, Average, or Low) × 4 (Problem Size: Very Small, Small, Large, or Very Large) × 2 (Instructional Bias: Speed or Accuracy)

		<i>F</i> Values for Each Reported Procedure			
Effect	<i>df</i>	Retrieval	Derived Facts	Guessed	Counting/ Addition
Between-Subjects					
Skill	2	3.65**	4.30**	1.98	0.13
<i>MS</i> _e	29	(1,355.7)	(435.4)	(376.9)	(180.3)
Within-Subjects					
Problem size (PS)	3	21.23***	12.73***	11.50***	3.23**
PS × skill	6	4.82***	1.94	1.36	1.58
<i>MS</i> _e	87	(223.4)	(133.6)	(91.56)	(64.4)
Instructional bias	1	0.29	9.97***	6.73**	0.39
Instructional bias × skill	2	0.60	4.13**	3.47**	1.42
<i>MS</i> _e	29	(75.8)	(43.6)	(99.6)	(16.5)
PS × instructional bias	3	1.29	2.32*	4.27***	2.98**
PS × instructional bias × skill	6	1.13	2.83**	2.55**	0.85
<i>MS</i> _e	87	(38.0)	(28.7)	(20.74)	(8.14)

p* < .09. *p* < .05. ****p* < .01

Figure 2. Percentages of reported use of retrieval across skill groups by problem size (very small, small, large, or very large) and instructional bias (speed or accuracy). The 95% confidence intervals are based on the MS_e for the three-way interaction of skill, problem size, and bias.

racy instructions (6% vs. 2%) but fewer derived facts (5% vs. 7%). This tradeoff between guessing and derived facts as a function of speed and accuracy requirements is consistent with the data in that derived fact procedures were slower but more accurate than guessing (see Table 4).

For all three nonretrieval procedures, problem size and instructional bias interacted (although only marginally for derived facts). These interactions were further qualified for guessing and derived fact solution procedures by the three-way interactions of problem size, instructional bias, and skill. As is shown in Figure 3, reports of derived facts increased more across problem size under accuracy instructions than under speed instructions. In contrast, reports of guessing increased more across problem size under speed instructions than under accuracy instructions. For counting and addition procedures, reported use generally decreased with problem size, but the pattern was slightly different under speed instructions than under accuracy instructions. In general, these data indicate that the participants' self-reports of procedures other than retrieval were sensitive to instructional biases. Furthermore, the self-reports corresponded to the latency and error data. With an accuracy bias, the participants relied more on derived fact procedures, whereas with a speed bias they were more likely to guess. Counting and addition procedures were used more on smaller problems when speed was emphasized and somewhat more on larger problems when accuracy was required. Thus, the participants responded in sensible ways to the instructional demands of the task.

Skill was also related to the participants' self-reports. The low-skill participants were most likely to report

using derived facts (11%), as compared with the high- (5%) or the average-skill (2%) participants. Moreover, the high- and the average-skill participants reported similar use of derived facts in speed and accuracy conditions (4% vs. 5% for high skill; 2% vs. 3% for average skill). In contrast, the low-skill participants reported using derived facts less frequently under speed instructions than under accuracy instructions (8% vs. 14%). These patterns were qualified by the interaction of skill, problem size, and instructional bias, as is shown in Figure 3. The high-skill participants reported more derived facts under accuracy bias than under speed bias on large problems only. Interestingly, the high-skill participants reported using more derived facts than did the average-skill participants. Thus, the similarity in the reports of retrieval between the high- and the average-skill groups masked differences in their use of other procedures. The average-skill participants reported significantly more derived facts under accuracy bias than under speed bias on the very large problems only and generally reported relatively few of these procedures. In contrast, the low-skill participants reported these procedures frequently. Furthermore, they reported using significantly more derived fact procedures under accuracy instructions on small, large, and very large problems, as compared with the speed bias. Thus, their reports of derived facts mirrored their patterns of latencies and errors.

The high-skill participants infrequently reported guessing (1% vs. 2% for the speed and the accuracy conditions). The average- and the low-skill participants guessed more frequently in speed conditions than in accuracy conditions (5% vs. 2% for average skill; 11% vs.

Figure 3. Percentages of reported use of counting and addition procedures (e.g., solving 3×5 as $5, 10, 15$), derived fact procedures (e.g., solving 8×9 as $8 \times 10 - 8$), and guessing across skill groups under speed and accuracy bias conditions. The 95% confidence intervals were calculated on the basis of the MS_e for each three-way interaction.

3% for low skill). As is shown in Figure 3, patterns for guessing mirrored those for derived facts. The high-skill participants were somewhat more likely to guess under accuracy requirements than under speed requirements on the very large problems. In contrast, the average- and the low-skill participants were more likely to report guessing when speed was emphasized over accuracy. For the average-skill participants, the difference across instructional biases was significant only on the very large problems (consistent with patterns of latencies). However, for the low-skill participants, guessing was reported more frequently in the speeded condition than in the accuracy condition for all sizes of problems.

Errors. The mean percentages of errors as a function of procedure reports for each participant on each procedure

are shown in Table 4 (averaging across problem size, because the number of errors in any given category was quite small). Errors were most frequent when the participants reported that they had guessed, although the majority of errors were made on retrieval trials, since that was the most commonly reported procedure. Comparisons were made across speed and accuracy conditions for each procedure. The participants made more errors on guesses with speed instructions than with accuracy instructions [34% vs. 16%; $t(31) = 3.13$, $SE = 5.64$]. Errors on retrieval trials were also more frequent with speed instructions than with accuracy instructions [5.3% vs. 2.7%; $t(31) = 3.29$, $SE = 0.80$]. Errors were similar across speed and accuracy instructions for derived facts and for counting/addition procedures.

DISCUSSION

Research on mental arithmetic has been influenced substantially by the claim that people use a variety of different solution procedures on single-digit problems (Campbell & Fugelsang, 2001; Campbell & Timm, 2000; Campbell & Xue, 2001). The present research provides further support for that view and adds to the literature by showing that instructional demands and individual differences can influence selection of solution procedures. As was predicted, individuals who were relatively slow and error-prone on arithmetic problems were most likely to be influenced by task demands and instructional requirements. Low-skill participants used a greater variety of procedures than did other individuals and were more likely to change their selection of non-retrieval procedures as a function of instructional emphasis. Asking these individuals to describe their solution procedures resulted in more accurate performance, but also in dramatically slower latencies. Thus, the results of this experiment indicate that reactivity to self-report requirements is selective. Furthermore, the correspondence between participants' self-reports and patterns of latencies and errors indicates that their self-reports accurately reflected their behavior.

Siegler (1987) demonstrated that averaging across different procedures could result in a misleading picture of performance (see also Haider & Frensch, 2002). We contend that averaging across skill levels in simple arithmetic tasks may also result in misleading conclusions. The problem size effect is greatly exacerbated when less skilled individuals are included (see Figure 1), especially when those individuals are using a high percentage of procedures other than direct retrieval (see also Campbell & Xue, 2001). Furthermore, these individuals are more likely to use slow (but accurate) procedures when given instructions to respond accurately than in speeded conditions, indicating that conclusions about performance should be tempered by the instructional context of the experiment.

Although the requirement to provide self-reports appeared to dramatically affect the behavior of low-skill participants, it is not appropriate to conclude that a speeded, silent control condition is the most veridical situation for collecting data from these individuals. In contrast to predictions that speed instructions would increase the use of retrieval, the individuals in the present research did not show such an effect. Instead, the combination of accuracy instructions and self-report requirements functioned as a strong manipulation of accuracy for the low-skill group (as is shown in Figure 1). This is a dramatic finding, because for low-skill participants in the silent control condition, accuracy instructions did not result in fewer errors than did speed instructions. For high- and average-skill participants, most trials probably reflect retrieval processes, and thus the nonretrieval trials exert relatively little impact on the latencies and patterns of errors. For low-skill participants,

however, collapsing across retrieval and nonretrieval trials may result either in an exacerbation of the problem size effect (if the participants are solving the problems accurately) or in an attenuation of the effect (if the participants are not solving the problems accurately). Thus, models of retrieval processes should be based only on data for which participants have used retrieval on the majority of trials.

Comparisons With Other Studies Using Self-Reports

Is the pattern of procedure use found in this study similar to those reported for multiplication problems in earlier studies? In the present study, reports of retrieval decreased with problem size, consistent with the results reported for multiplication by Campbell and Xue (2001), LeFevre, Bisanz, et al. (1996), and Hecht (1999). In contrast, the participants in Kirk and Ashcraft (2001, Experiment 2) reported using *more* retrieval as problem size increased (see their Figure 4, p. 169). In the present study, only the counting and the addition procedures were more common among smaller problems than among larger problems, and counting and addition procedures represented a very small portion of the overall procedure use (less than 3% overall). The participants in Kirk and Ashcraft (Experiment 2) reported using counting and addition procedures on 2.4% of the trials in the retrieval bias condition (similar to the present research), as compared with 18% and 23% in the self-report and strategy bias conditions, respectively. Thus, the instructions used by Kirk and Ashcraft may have influenced participants differently than did those used in previous research and in the present experiment.

In other respects, however, the participants in Kirk and Ashcraft's (2001) silent control and replication conditions showed performance similar to that found in the present research for the silent control and the self-report conditions (the latter was termed the replication condition by Kirk and Ashcraft). As in the present research, Kirk and Ashcraft found that latencies were very similar in the silent control and the self-report conditions. Furthermore, the participants made more errors in the silent control condition than in the self-report condition (i.e., 5.5% vs. 3.5%). In summary, the patterns across silent control and self-report conditions found in the present research were comparable to those observed in other studies.

Conclusions

Although it is clearly very important to develop methods other than self-reports for assessing the procedures that people use in arithmetic tasks (Penner-Wilger, Leth-Steensen, & LeFevre, 2002; Siegler & Lemaire, 1997), the present research indicates that it is not necessary to discount the accumulated evidence based on self-reports. First, high- and average-skill participants in the present research were not reactive to self-report requirements. Second, those who were reactive showed patterns

of behavior that were consistent with the self-reports. Thus, self-reports appeared to be veridical reflections of behavior for the individuals in the present research. The judicious use of self-report instructions in research on mental arithmetic is likely to provide useful information for developing comprehensive models.

The results of the present research clearly show how important it is to consider individual differences in mental arithmetic (LeFevre & Kulak, 1994). Other research supports this conclusion for mathematical cognition and for other complex cognitive tasks. Ashcraft and Kirk (2001) explored how individual differences in math anxiety affected performance in an arithmetic task. Schunn and Reder (2001) proposed a new measure of individual differences, adaptive selection of strategies, and showed how such individual differences were important in understanding performance in an air traffic control task. Thus, models of cognitive processing must include parameters that index individual differences in how people perform complex mental activities.

Another important conclusion from the present research is that the common experimental procedure of presenting simple arithmetic problems and asking participants to respond quickly does not necessarily provide the definitive perspective on solutions to these problems. If performance varies depending on individual and situational factors, a complete understanding of how people solve simple arithmetic will need to include a description of these factors and their effects. In addition to considering the impact of individual and situational factors, researchers must consider the consequences of averaging response latencies and errors across potentially different types of cognitive procedures. For example, Siegler (1987) demonstrated the pitfalls of averaging performance across different procedures. However, less attention has been paid to the related problems that may occur if performance is averaged across individuals who use very different combinations of solution procedures (Haider & Frensch, 2002; LeFevre, Sadesky, & Bisanz, 1996).

Finally, some researchers might find it surprising that adults continue to use relatively complex procedures, such as derived facts, on certain problems. The present research, however, suggests that use of such procedures may allow participants to moderate their performance—in particular, to respond accurately on certain problems on which they might otherwise make an error. The decision to maximize accuracy rather than speed on arithmetic problems may reflect real-world considerations: In many situations, such as answering the skill-testing question in a lottery, it might be much more important to get the correct answer than to respond quickly.

REFERENCES

- ASHCRAFT, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75-106.
- ASHCRAFT, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, **1**, 3-34.
- ASHCRAFT, M. H., & KIRK, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, **130**, 224-237.
- BAROODY, A. J. (1984). An evaluation of evidence supporting fact-retrieval models. *Learning & Individual Differences*, **6**, 148-156.
- CAMPBELL, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, **53**, 1-44.
- CAMPBELL, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, **1**, 121-164.
- CAMPBELL, J. I. D., & FUGELSANG, J. (2001). Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition*, **80**, B21-B30.
- CAMPBELL, J. I. D., & GRAHAM, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, **39**, 338-366.
- CAMPBELL, J. I. D., & OLIPHANT, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills* (pp. 331-364). Amsterdam: Elsevier.
- CAMPBELL, J. I. D., & TIMM, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review*, **7**, 692-699.
- CAMPBELL, J. I. D., & XUE, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, **130**, 299-315.
- COONEY, J. B., & LADD, S. F. (1992). The influence of verbal protocol methods on children's mental computation. *Learning & Individual Differences*, **4**, 237-257.
- FRENCH, J. W., EKSTROM, R. B., & PRICE, I. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
- GEARY, D. C., FRENCH, P. A., & WILEY, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed of processing differences in younger and older adults. *Psychology & Aging*, **8**, 242-256.
- GEARY, D. C., & WILEY, J. G. (1991). Cognitive addition: Strategy choice and speed of processing differences in young and elderly adults. *Psychology & Aging*, **6**, 474-483.
- HAIDER, H., & FRENCH, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 392-406.
- HECHT, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, **27**, 1097-1107.
- KIRK, E. P., & ASHCRAFT, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 157-175.
- LEFEVRE, J.-A., BISANZ, J., DALEY, K. E., BUFFONE, L., GREENHAM, S. L., & SADESKY, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, **125**, 284-306.
- LEFEVRE, J.-A., & KULAK, A. G. (1994). Individual differences in the obligatory activation of addition facts. *Memory & Cognition*, **22**, 188-200.
- LEFEVRE, J.-A., KULAK, A. G., & HEYMANS, S. L. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science*, **24**, 276-289.
- LEFEVRE, J.-A., & LIU, J. (1997). Numerical cognition: Single-digit multiplication skills of adults from China and Canada. *Mathematical Cognition*, **3**, 31-62.
- LEFEVRE, J.-A., & MORRIS, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition*, **27**, 803-812.
- LEFEVRE, J.-A., SADESKY, G. S., & BISANZ, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 216-230.
- LEFEVRE, J.-A., SMITH-CHANT, B. L., HISCOCK, K., DALEY, K. E., & MORRIS, J. (2003). Young adults' strategic choices in simple arithmetic: Implications for the development of mathematical representa-

- tions. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Recent research and theory* (pp. 203-228). Hillsdale, NJ: Erlbaum.
- LEMAIRE, P., & SIEGLER, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, **124**, 83-97.
- LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.
- PENNER-WILGER, M., LETH-STEENSEN, C., & LEFEVRE, J.-A. (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory & Cognition*, **30**, 1160-1167.
- RUSSO, J. E., JOHNSON, E. J., & STEPHENS, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, **17**, 759-769.
- SCHUNN, C. D., & REDER, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, **130**, 59-76.
- SIEGLER, R. S. (1987). The perils of averaging over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.
- SIEGLER, R. S. (1988a). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, **59**, 833-851.
- SIEGLER, R. S. (1988b). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, **117**, 258-275.
- SIEGLER, R. S., & JENKINS, E. A. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- SIEGLER, R. S., & LEMAIER, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, **126**, 71-92.
- SIEGLER, R. S., & SHIPLEY, E. (1995). Variation, selection and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modelling* (pp. 31-76). Hillsdale, NJ: Erlbaum.
- SIEGLER, R. S., & SHRAGER, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.
- SVENSON, O. (1985). Memory retrieval of answers of simple additions as reflected in response latencies. *Acta Psychologica*, **59**, 285-304.

NOTES

1. Due to an error, the problem 4×6 appeared only once (instead of twice) in one list, whereas 6×4 appeared three times in that list. In the other list, 9×6 was never shown, and the participants saw 9×2 four times in that list. All other problems were presented twice on each list.
2. Order of instructional bias (i.e., speed first vs. accuracy first) was counterbalanced across participants. When order was included in the ANOVA, there was a significant interaction with instructional bias. Although the participants were always faster in the speed condition than in the accuracy condition (1,289 vs. 1,396 msec for the speed-accuracy order; 989 vs. 1,310 msec for the accuracy-speed order), the difference was greater in the latter case—that is, when the speeded condition was second (CI = 83 msec). No other effects involving order were statistically significant.

(Manuscript received October 19, 2001;
revision accepted for publication October 15, 2002.)