Journal of Experimental Psychology:
Learning, Memory, and Cognition
1996, Vol. 22, No. 5, 1281-1295


0278-7393/96/$3.00

# Some Tests of an Identical Elements Model of Basic Arithmetic Skills

Timothy C. Rickard
National Institute of Neurological Disorders and Stroke

Lyle E. Bourne, Jr.
University of Colorado at Boulder

Two experiments tested an identical elements model of the organization of basic arithmetic skills (T. C. Rickard, A. F. Healy, & L. E. Bourne, 1994). This model assumes a distinct abstract representation for each unique combination of the basic elements (i.e., the operands and the required operation) of a problem. Participants practiced multiplication and division problems and were then tested on various altered versions of these problems. Experiment 1 confirmed the prediction of no positive transfer when the presented elements of a test problem do not exactly match those of a practice problem. Experiment 2 confirmed the prediction that there is complete transfer when the elements of the test problem match exactly with those of a practice problem. Experiment 2 also confirmed that there is both perceptually specific and nonspecific speed-up with practice. Implications for number processing and arithmetic are discussed.

Recent empirical investigations of numerical cognition have demonstrated that skilled (i.e., adult) performance on basic arithmetic (e.g., 3 × 6 = ?) typically reflects retrieval of discrete facts from memory rather than execution of more generic calculational algorithms (for a review see Ashcraft, 1992). This finding raises several fundamental questions about the structure and organization of this knowledge in memory. For example, are arithmetic problems that are presented in different modalities or formats (e.g., 4 × 7 and four times seven) best seen as accessing a single semantic memory structure, or separate, modality-specific memory structures? Do complementary operand orders of a problem (e.g., 6 × 8 and 8 × 6), or complementary problems involving two different operations (e.g., 3 × 9 = _ and 3 × _ = 27), or related problems within a noncommutative operation (7 × _ = 35 and 5 × _ = 35), access the same, or different, memory structures? How do other, somewhat more subtle differences in problem presentation, such as vertical versus horizontal visual format, or variations of the symbol used to denote an arithmetic operation, affect processing? Although all of these questions focus on arithmetic, there are clearly analogs in other domains. In the present study we tested an identical elements model proposed by Rickard et al. (1994), which provides a relatively straightforward and concise starting point from which to investigate these issues.

The identical elements model incorporates three basic assumptions. First, for simplicity the model assumes distinct and sequential perceptual, cognitive, and motor stages of performance. The cognitive stage involves access to semantic representations that are independent of the modality-specific representations within the perceptual stage. Second, it assumes that answer retrieval occurs exclusively within the cognitive stage. These assumptions are related to those of a modular number processing theory proposed by McCloskey, Caramazza, and Basili (1985). There is empirical support for this three-stage approach (e.g., McCloskey, 1992), although there is also support for other contrasting number processing frameworks (e.g., Campbell & Clark, 1992; Deheane, 1992). The third assumption is that representation of each arithmetic fact within the cognitive stage can be fully characterized in terms of its three essential constituent elements: one element corresponding to each of the two presented numbers (e.g., 4 and 7 for the problem 4 × 7), and one for the operation formally required (e.g., "multiply"). Note that "the operation formally required" refers to the operation required in the mathematical sense, rather than to the arithmetic symbol present in the problem. For example, the answer to 4 × _ = 28 requires division. For commutative operations, the order of the numbers is not represented. Thus, for example, the two operand orders of a multiplication problem map on to the same unitary representation within the cognitive stage. For noncommutative operations, the order is preserved (thus, 20 ÷ 4 and 4 ÷ 20 would be represented uniquely). None of the remaining format or modality-specific features are represented within this stage.

The identical elements models implies straightforward answers to each of the questions stated above. For example, a problem presented in arabic format, such as 4 × 7, and the same problem presented in a written verbal format, such as *four times seven*, will access the exact same semantic memory "chunk" within the cognitive stage. Similarly, multiplication problems that differ only in operand order, such as 3 × 8 and 8 × 3, will access the same memory chunk. In summary, any problems that differ only with respect to the format of presentation or the relational characteristics among the elements (e.g., operand order, horizontal vs. vertical presentation, or variations in the arithmetic symbol used) will access the same chunk. In contrast, problems that differ with respect

Timothy C. Rickard, Cognitive Neurocience Section, Medical Neurology Branch, National Institute of Neurological Disorders and Stroke; Lyle E. Bourne, Jr., Department of Psychology, University of Colorado at Boulder.

This research was supported in part by Army Research Institute Contract MDA903-90-K-0066 to the University of Colorado. We thank Alice Healy for her thoughtful comments concerning this research.

Correspondence concerning this article should be addressed to Timothy C. Rickard, Cognitive Neuroscience Section, Medical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Building 10, Room 5S209, 10 Center Drive MSC1440, Bethesda, Maryland 20892-1440.

to one or more elements will access completely different memory chunks. Thus, 4 × 7 and 7 × 6 will access completely different memory chunks. A somewhat less intuitive prediction is that complementary problems from two operations (e.g., 4 × 7 = _ and 4 × _ = 28) and related problems within a noncommutative operation (e.g., 28 = _ × 4 and 28 = _ × 7) access completely different memory chunks.

Note that the model is not necessarily intended to apply to the initial stages of skill acquisition. Rather, the model is designed to account for the asymptotic organization of memory after extended experience. In essence the model makes predictions about what sorts of semantic generalizations will and will not occur within the long-term memory system that supports "expert" arithmetic fact retrieval. Empirical verification of the model (or of a competing model) would suggest principles of memory organization that may have implications for the storage and retrieval of facts in a variety of domains. Also, once a clear and empirically powerful model of memory organization in skilled arithmetic is identified, it may suggest productive new questions regarding the skill acquisition mechanisms that mediate the transformation from novice to expert performance in this and perhaps other domains.

Rickard et al.'s (1994) Experiment 1 provided the initial data suggesting the identical elements model, and Experiment 2 of that study provided some confirming evidence. In Experiment 1, participants practiced over three 1 hr sessions on a set of basic multiplication and division problems (e.g., 4 × 7 = 28 and 9 × _ = 45) and were then tested on the same problems seen during practice (*no change* problems) and also on *operand order change* problems, *operation change* problems, and *operation plus order change* problems, as shown in Table 1 (ignoring the "new problems" entry for now). Problems at test were analyzed separately by operation (multiplication or division), and thus Table 1 lists the test conditions for the two operations separately. For multiplication, error rates were about 3 times greater and response times (RTs) were about 350 ms slower for the operation and operation plus order change conditions than for the no change and operand order change conditions. For division, the pattern of results was substantially different. There were no reliable differences among the operand order, operation, and operation plus operand order conditions, and performance in each of these conditions was substantially

Table 1
*Test Conditions for an Example Problem at Test in Experiment 1*

| Test condition | Practice | Test |
|---|---|---|
| Test condition | | |
| No change | 4 × 7 = ___ | 4 × 7 = ___ |
| Order change | 7 × 4 = ___ | 4 × 7 = ___ |
| Operation change | 4 × __ = 28 | 4 × 7 = ___ |
| Order and operation change | 7 × __ = 28 | 4 × 7 = ___ |
| New problems | | 4 × 7 = ___ |
| Division | | |
| No change | 4 × __ = 28 | 4 × __ = 28 |
| Order change | 7 × __ = 28 | 4 × __ = 28 |
| Operation change | 4 × 7 = ___ | 4 × __ = 28 |
| Order and operation change | 7 × 4 = ___ | 4 × __ = 28 |
| New problems | | 4 × __ = 28 |

slower (by about 400 ms) and about 3 times more error prone than was performance in the no change condition. Each of these results is consistent with the model. In the no change condition for both operations, as well as in the operand order change condition for multiplication, the elements of the problems as defined by the model were exactly the same as the elements of a problem seen during practice, and thus the model predicts substantial transfer of any speed-up that occurred during practice. In all of the remaining test conditions, the elements of the problems did not match exactly with those of a practice problem, and thus no transfer within the cognitive stage would be expected.

Rickard et al.'s (1994) Experiment 2 employed a similar practice–transfer design. However, at test changes were made in the operation to be performed (as in Experiment 1) as well in the operation symbol presented (e.g., _ = 4 × 7 during practice, and _ ÷ 4 = 7 during test). As predicted by the model, a symbol change at test had only a negligible impact on RTs and error rates, whereas an operation change again had a very substantial impact.

As a shorthand we can refer to test conditions in which elements of the problem match exactly with those of a practice problem as *noncritical change* conditions, and to test conditions in which problem elements do not match exactly as *critical change* conditions. The identical elements model predicts substantial positive transfer of training to test problems with noncritical changes, but no transfer of training to test problems with critical changes, with the possible exception discussed below of general transfer effects originating in the perceptual stage, motor stage, or both.

Although Rickard et al.'s (1994) results were generally consistent with the model, there were two salient patterns in the data that they did not directly predict. First, a comparison of performance at the beginning of practice with performance on problems with critical changes at test revealed, for both experiments, reliable positive transfer measured by RTs to problems with critical changes at test. This positive transfer can be interpreted generally in two ways: (a) The model may simply be incorrect, or (b) there may have been speed-up in general processes outside of the cognitive stage (such as speed-up in execution of the motor responses required for the number keypad response mode that was used, or speed-up due to increased familiarity with the general format in which problems were presented) during practice (see Rickard et al., 1994, for a more detailed discussion).

Second, for multiplication problems in Experiment 1, RTs in the operand order change were reliably slower (by about 80 ms) than RTs in the no change condition (see also Fendrich, Healy, & Bourne, 1993, for an earlier documentation of this effect). An analogous effect of slightly slower performance in the symbol change relative to the no change condition was observed in Experiment 2. The model does not directly predict these effects, and the only way it can potentially account for them is to assume that a perceptual processing advantage accrued for the practiced operand order (in Experiment 1) or the practiced symbol (in Experiment 2), resulting in faster access for no change problems to the single underlying memory structure within the cognitive stage.

The current studies were designed in part to determine whether the identical elements account of the two results discussed above is correct. Experiment 1 tested the prediction of no positive transfer with a change in operation from practice to test, using a methodology that provided much improved control over potential general transfer effects. Experiment 2 examined whether the RT advantage for no change over operand order change multiplication problems (Fendrich et al., 1993; Rickard et al., 1994) reflects solely a perceptual advantage for no change problems, using a methodology that factored out any perceptually based differences between these conditions. In addition, the design of Experiment 2 allowed us to explore the relative extent to which speed-up with practice is or is not directly tied to the format in which the problems are practiced.

## Experiment 1

This experiment extends Rickard et al.'s (1994) Experiment 1 by adding a *new problems* condition at test for both multiplication and division (see Table 1). These new problems were not seen during practice in either operation. The identical elements model makes the following predictions. First, there should be reliably lower error rates and reduced RTs at test for noncritical change compared with critical change problems. Second, because the various general perceptual and motor transfer factors discussed by Rickard et al. (1994) are equated among the critical change test conditions in this experiment, the model makes the prediction that performance among the critical change conditions will not differ. Other results already in the literature allow us to make additional inferences if this second prediction is confirmed. Specifically, Campbell (1987), and Pauli, Lutzenberger, Birbaumer, Rickard, and Bourne (1994) have shown that performance on new (unpracticed) problems at test is no faster than performance on the first block of practice, using experimental designs that likely minimize general transfer effects (e.g., Campbell used a voice-key response modality that likely minimized or eliminated general speed-up in the motor responses during practice). These results are themselves consistent with the model. In addition, they allow a finding of no performance differences between the new problems condition and the other critical change conditions to be validly interpreted as a finding of no positive transfer to any of the critical change conditions.

One additional issue was explored in this experiment. Rickard et al. (1994) observed an RT advantage for multiplication over division throughout 40 blocks of practice. They suggested that more frequent exposure to multiplication might be responsible for this effect. However, there was also an interaction between operation and amount of practice in these data, such that the performance difference between the two operations decreased reliably with practice. This result suggests that with enough practice, division performance might overtake multiplication performance, an effect that would be difficult to reconcile with a pure frequency account of the multiplication advantage. In the current experiment, there were 90 blocks of practice on both multiplication and division problems, thus allowing for a more thorough investigation of relative multiplication and division performance.

### *Method*

#### *Participants*

Twenty-four students in an introductory psychology class participated in the experiment for course credit.

#### *Apparatus*

Participants were tested on Zenith Data Systems personal computers, programmed with the Micro Experimental Language (MEL) software (Schneider, 1988).

#### *Materials and Procedure*

Participants received three sessions of practice (30 blocks per session, 16 problems per block) and were then tested on a fourth session. Practice sessions were given Monday, Wednesday, and Friday of one week, and the test session was given the following Monday. All sessions lasted from 30 to 45 min. Twelve practice sets (2 participants per set) were created according to the following specifications. Twenty-four arabic arithmetic problems between $2 \times 5 = 10$ and $8 \times 9 = 72$ were selected as the stimulus items. These 24 problems were divided into three subgroups of 8, with problems in each subgroup being roughly equated on difficulty. These problem groups are shown in Table 2. Three master sets of 16 problems were then created, one corresponding to each of the possible pairings of the three subgroups. Within each master set, one subgroup was assigned as multiplication problems (e.g., $6 \times 9 = \_$), and the other as division problems (e.g., $6 \times \_ = 54$). The multiplication symbol, ×, was used for all problems. Half of the problems within each subgroup of each master set had ascending operand order, and half had descending operand order. These three master sets constituted three of the actual practice sets. Three additional practice sets were then derived from each of the master sets by reversing the operand order, reversing the operation, or reversing both the operand order and the operation, of all problems within the set. Thus, 4 practice sets were derived from each of the three master sets, yielding a total of 12 practice sets of 16 problems. Participants saw problems one at a time presented in the middle of the screen and entered the one- or two-digit answer using the number keypad. Feedback was given only following problems on which a participant's answer was incorrect. The computer automatically proceeded to the next problem after the participant entered the last digit of the answer, or after the accuracy feedback. Problems were presented randomly within each block. At test, each participant saw each of the 24 problems in both operand orders and in both operations once during each of five blocks (thus there were 96 problems per block at test). All other details of problem presentation during test were the same as those for the practice problems. Note that for convenience we use the term *operand* to refer to the single-digit numbers for both

Table 2
*The Three Groups of Problems From Which the Master Sets for Experiment 1 Were Constructed*

| Set 1 | Set 2 | Set 3 |
|---|---|---|
| 3 × 5 = 15 | 5 × 4 = 20 | 3 × 7 = 21 |
| 9 × 2 = 18 | 2 × 6 = 12 | 4 × 3 = 12 |
| 4 × 6 = 24 | 6 × 5 = 30 | 2 × 8 = 16 |
| 7 × 5 = 35 | 3 × 6 = 18 | 8 × 5 = 40 |
| 6 × 7 = 42 | 8 × 7 = 56 | 5 × 9 = 45 |
| 8 × 9 = 72 | 9 × 3 = 27 | 7 × 4 = 28 |
| 6 × 8 = 48 | 4 × 9 = 36 | 8 × 3 = 24 |
| 9 × 7 = 63 | 8 × 4 = 32 | 9 × 6 = 54 |

multiplication and division problems, and *product* to refer to the double-digit number. For multiplication, the product is the answer, and for division one of the operands is the answer.

## *Results*

Unless otherwise indicated, all statistical results are reliable at the .05 level. Error analyses were performed on the raw error proportions. Secondary analyses on the arcsine transformed error proportions yielded an equivalent set of reliable effects in all instances. RT analyses were performed on the log transformed initiate RT (the interval between the onset of the problem on the computer screen and the pressing of the first digit of the answer) and were limited to correctly solved problems. Initiate RT is strongly correlated with total RT, the interval between the onset of the problem, and the pressing of the second digit of the answer (Fendrich et al., 1993). All reported test analyses were performed on data collapsed across all problems within each condition. Supplementary analyses performed on problems split into easy and difficult categories (defined on the basis of normative data collected by Campbell & Graham, 1985) revealed the same sets of reliable and nonreliable results as did the overall analyses.

### *Practice*

The overall error proportions for multiplication problems in Sessions 1, 2, and 3 were .0350, .0283, and .0283, respectively. Corresponding values for division problems were .0283, .0166, and .0163, respectively. Figure 1 shows log RTs for correctly solved problems plotted by log block and operation (multiplication or division). Each data point represents up to 192 observations, with data collapsed over participants and problems. (The frequencies did not always reach these limits because error trials were excluded form the RT analyses.) A within-subjects regression analysis with log block as a continuous variable and operation (multiplication vs. division) as a categorical variable was performed on the mean log RT of each block. The results are shown graphically in Figure 1. The overall correlation for this analysis was .86. There was a reliable overall effect of log block, $F(1, 23) = 291.00$, and an interaction between log block and operation, $F(1, 23) = 25.1$. The advantage for multiplication (205 ms) on the first block of practice was reliable, $F(1, 23) = 6.10$, but there was a trend toward a division advantage (46 ms) on the last block of practice, $F(1, 23) = 2.54, p = .12$. Although the log–log power function fits were quite good for most of the data, there were clear deviations from linearity for the first block of each session (Blocks 1, 31, and 61), as evident in the figure. The data at the individual trial level revealed that this effect primarily reflects a roughly 1,000 ms slower response latency on the very first trial of the first block of each session. Participants apparently required additional time on the first trail of each session to orient to the task. Exclusion of these practice blocks from the regression fit produced only negligible differences in the results.

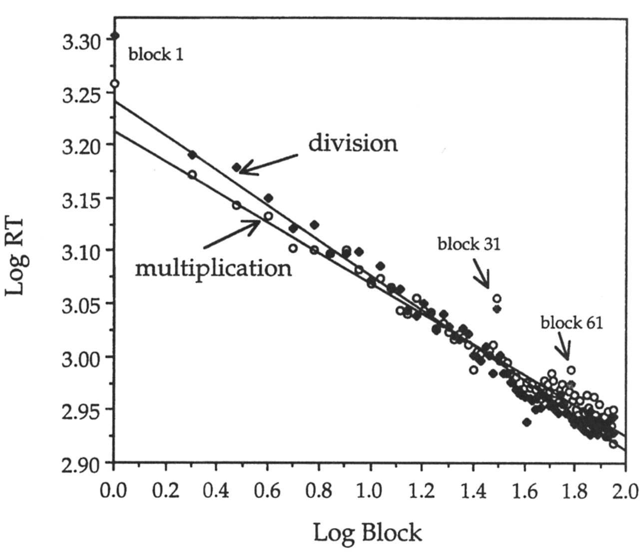


*Figure 1.* Log RT for correctly solved problems in Experiment 1 plotted as a function of log block and operation (multiplication or division). RT = response time.

### *Test: Multiplication*

*Error analyses.* Figure 2a shows the error proportions for multiplication problems in each of the five test conditions, collapsed across block. A within-subjects analysis of variance (ANOVA) with one five-level variable of test condition (no change, order change, operation change, operation and order change, and new problems) was performed on these proportions. The overall effect of test condition was reliable, $F(4, 23) = 8.7$. Four planned orthogonal contrasts were also evaluated. The first contrast, comparing the no change and operand order change conditions with the remaining three conditions, was strongly reliable, $F(1, 23) = 34.6$, confirming a basic prediction of the identical elements model that performance should be substantially better for problems in noncritical compared with critical change conditions. The second contrast, comparing the no change condition with the operand order change condition; the third contrast, comparing the operation and operation plus operand order change condition with the new problems condition; and the fourth contrast, comparing the operation change condition with the operation plus operand order change conditions, were not significant (all $Fs < 1.0$).

*RT analyses.* The anti-log of the mean log initiate RT for multiplication problems is plotted by block and test condition in Figure 2b. A 5 × 5 within-subjects ANOVA with variables of block (1–5) and test condition was performed on the log RTs. There were reliable effects of both block, $F(4, 23) = 17.1$, and test condition, $F(4, 23) = 96.1$. There was also a reliable interaction of block and test condition, $F(16, 23) = 4.6$, reflecting speed-up across the blocks of test that was essentially limited to problems in the critical change conditions (i.e., the operation change, operation plus order change, and new problems conditions). This interaction was not reliable when limited to the critical change conditions. Four planned orthogonal contrasts, identical to those discussed for the error analysis, were evaluated. The first of these, comparing performance in the noncritical and critical change conditions, was strongly reliable, $F(1, 23) = 373.7$. The second, comparing the no

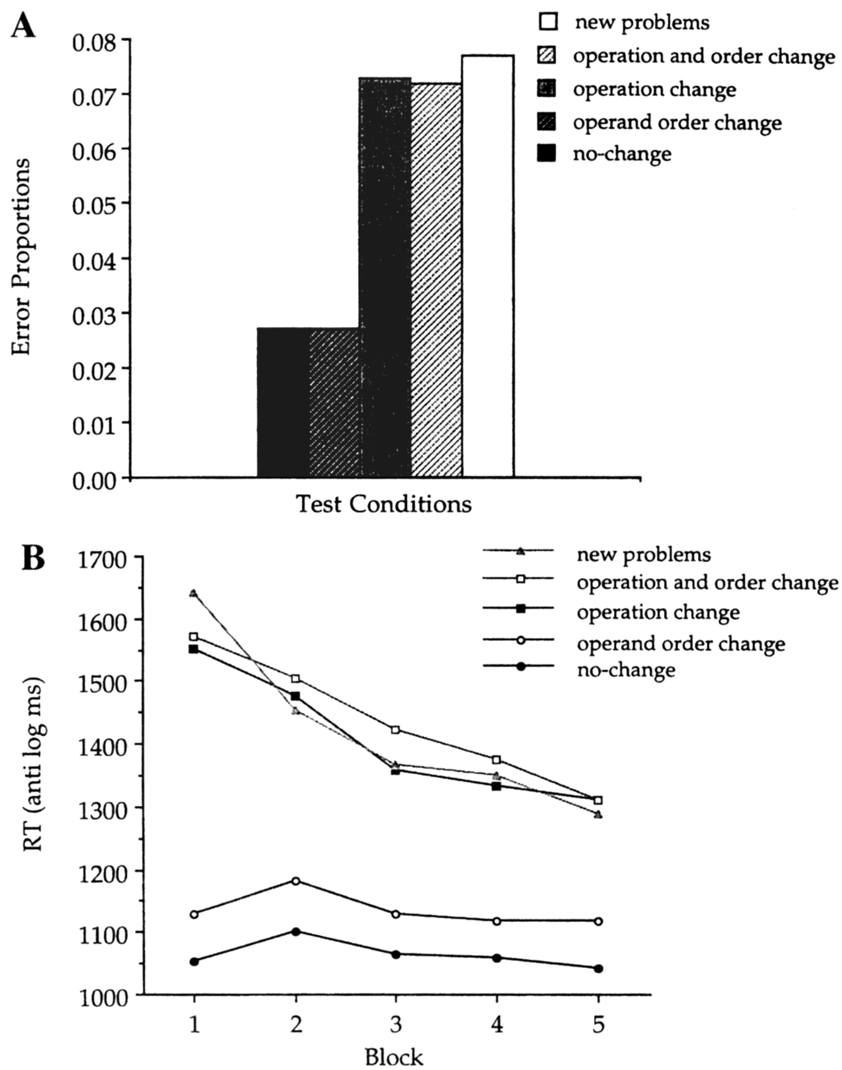


*Figure 2.* A: Error rates and B: anti-log RTs for multiplication problems in Experiment 1. Error data are collapsed across test blocks. RT = response time.

change condition with the operand order change condition, was also reliable, $F(1, 23) = 9.76$. The third and fourth contrasts, comparing performance among problems in the critical change conditions (and identical to the third and fourth contrasts performed for the error analyses), were not reliable ($Fs < 1.12$). Thus, for both error and RT data, there were large differences between noncritical and critical change conditions, but there was no evidence of any performance differences among the critical change conditions.

### Test: Division

*Error analyses.* Figure 3a shows the error proportions for division problems in each of the five test conditions, collapsed across block. The overall effect of test condition was reliable in a within-subjects ANOVA, $F(4, 23) = 6.3$. The first planned orthogonal contrast, comparing the no change condition with the remaining four conditions, was also reliable, $F(1, 23) = 23.0$. This result confirms the basic prediction of better performance in the noncritical change condition than in the critical change conditions. The remaining contrasts, comparing operand order change, operation change, and operation plus operand order change conditions with the new problems condition, $F(2, 23) = 1.4$; comparing the operand order change condition with the operation and operation plus operand order change conditions, $F(1, 23) = 0.50$; and comparing the operation change condition with the operation plus operand order change condition, $F(1, 23) = 0.09$, were not reliable. Thus, there was no statistical evidence of any difference among conditions representing critical changes from practice to test.

*RT analyses.* The anti-log of the mean log initiate RT for correctly solved division problems is shown in Figure 3b. A 5 × 5 within-subjects ANOVA with variables of block (1–5) and test condition revealed reliable main effects of both block, $F(4, 23) = 11.0$, and test condition, $F(4, 23) = 64.5$. The interaction

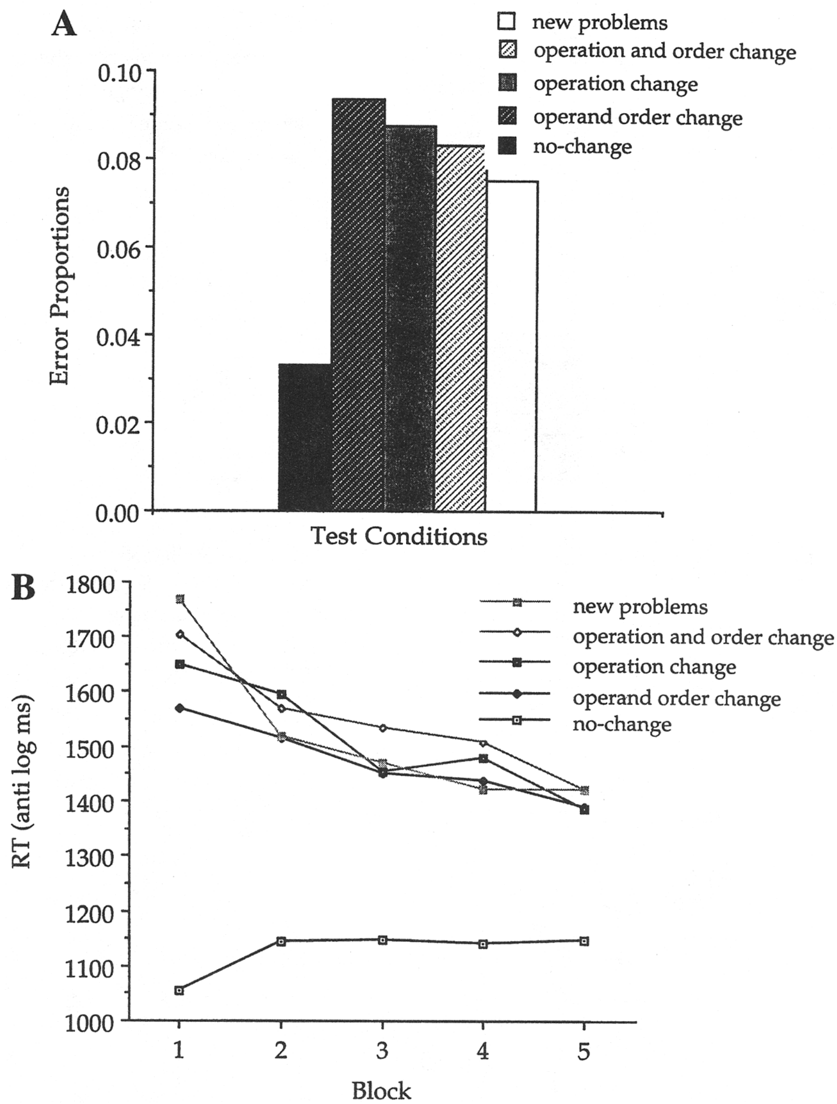


*Figure 3.* A: Error rates and B: anti-log RTs for division problems in Experiment 1. Error data are collapsed across test blocks. RT = response time.

of block and test condition was also reliable, $F(16, 23) = 5.4$, reflecting speed-up during the test that was limited to the critical change conditions. This interaction was not reliable when limited to the critical change conditions. Four planned orthogonal contrasts of the test condition variable (identical to those performed in the division error analysis) were evaluated. The first of these, comparing the no change condition with the four critical change conditions, was strongly reliable, $F(1, 23) = 253.6$. The second, comparing the operand order change, operation change, and operation plus operand order change conditions with the new problems condition, and the third, comparing the operation change condition with the operation plus operand order change condition, were not reliable ($F$s $< 1.0$). The fourth contrast, comparing the operand order change condition with the operation and operation plus operand order change condition, approached significance, $F(1, 23) = 3.44$, $p = .067$.

## *Discussion*

### *Practice*

The practice data replicated Rickard et al.'s (1994) findings of both an advantage for multiplication at the beginning of practice and a reduction in the multiplication advantage with practice. In addition, in the present experiment there was a crossover interaction such that, by the end of practice, division problems were solved faster than multiplication problems. The crossover suggests the possibility that two factors influence relative multiplication and division performance. First, a frequency advantage provides a plausible account for the faster multiplication RTs that we have now observed in two experiments at the beginning of practice. Note that the identical elements model, if correct, makes a frequency advantage for multiplication very likely. The two operand orders of a multiplication problem map onto a single underlying knowledge structure in the cognitive stage, whereas the operand orders of a division problem (as defined in this experiment and by Rickard et al., 1994) map onto completely different knowledge structures. It is also reasonable to assume that in everyday life, the two operand orders of a multiplication problem are encountered at least as often as the two operand orders of a division problem. Given these assumptions, there should be a frequency advantage for multiplication over division in the natural environment. During practice in this experiment, however, there was no such advantage for multiplication; only one operand order of each multiplication and division problem was presented, and multiplication and division problems were presented equally often. Thus, the advantage for multiplication over division would be expected to decrease throughout practice.

Frequency effects alone, however, cannot account for the finding that division problems appear to be solved faster by the end of practice. One possible account of the eventual division advantage appeals to the different combinatorial structure of the elements of the multiplication and division problems presented during practice. For multiplication, each operand was present in several problems. For division, each product was present in only one problem. Using the terminology of the well-established fan effect (Anderson, 1974), the product for division problems has an effective fan of 1 (each product is associated with only one answer) during practice, whereas both operands of each multiplication problem have fans of 3 or 4. High-fan items typically yield slower response times than low-fan items, and this factor may account for the faster division performance by the end of practice.

One way to test this hypothesis is to look for interference that the strong product–answer associations for division problems would be expected to produce when different problems are presented at test. For example, if strong product–answer associations form during practice they might interfere with performance on novel division problems at test that have the same product as a division problem presented during practice. All problems in the operand order change condition are candidates (e.g., $7 \times \_ = 63$ during practice; $9 \times \_ = 63$ during test). In addition, some problems in the remaining critical change conditions are candidates. Specifically, pairs of problems having the same product but different sets of operands (e.g., $4 \times \_ = 24$ and $6 \times \_ = 24$, vs. $3 \times \_ = 24$ and $8 \times \_ = 24$) were arranged in the design such that one member from each pair was seen during practice, and the others were seen at test. Thus, if $4 \times \_ = 24$ was seen during practice, the other three problems were seen at test. One of these problems in this example, $6 \times \_ = 24$, is the operand order change problem as discussed above. The other two would fall in the operation change, operation plus operand order change, or new problems condition. All of these problems would potentially be subject to the hypothesized product-based interference, which would be reflected in this case by an error response of 6.

A post hoc error analysis appears to confirm the product-based interference hypothesis. Participants made 116 errors involving division problems that could be subject to this product-based interference from practice. Of these, 57 (49%) were indeed practice-related errors; that is, they were the correct answer for the problem solved during practice that contained the same product. Assuming that the eight division answers used in the experiment constitute the set of possible answers to these problems, and noting that only one of these answers is the correct one, then there are seven different potential errors for each division problem. If each of these seven errors is equally likely to occur for each problem, then we would expect an error corresponding to the correct answer to the practice problem with the same product to occur only 14 % of the time. The actual percentage of product-based errors was roughly three and one-half times greater than would be expected by chance, and the observed percentage was greater than chance for 20 of the 23 participants who exhibited errors for these problems, an effect that was strongly reliable using a binomial sign test ($p < .001$).

One potential confound in this analysis is that for some of the problems subject to possible product-based interference from practice, a product-based error can not be differentiated from an operand naming error similar to the operand intrusion errors discussed by Campbell (1994). For example, the error 6 for the problem $6 \times \_ = 24$ might simply reflect an intrusion of the presented number 6 as the answer. However, two sources of evidence allow us to reject an operand intrusion account of this result. First, the proportion of operand matching errors on problems for which product-based errors could not occur was only .23. In contrast, as discussed above, the proportion of operand matching errors for problems on which product-based errors could occur was .49 on average over participants, and this proportion was greater than .23 for 20 of 23 participants (the lowest proportion among these participants was .21). Thus, at least some of these errors must reflect a factor other than operand intrusion, and the only apparent alternative is the hypothesized product-based interference effect. Further evidence for product-based interference at test is garnered from a comparison of division and multiplication RTs in the no change conditions. Whereas there was a trend toward slower RTs for division than for multiplication by the end of practice, the situation reversed for the no change conditions at test, with multiplication being reliably faster than division (1,064 and 1,127 ms, respectively), $F(1, 23) = 5.86$. This effect may reflect an overall adjustment in speed–accuracy criterion for division at test to control the errors caused by product-based interference.

In summary, development of an unusually strong product–answer association for division during practice appears to be a good candidate to explain both the superior performance on division by the end of practice and the interference effects for division during test. According to this account, both the facilitation and interference effects are critically dependent on the fact that only one operand order of each division problem was presented to participants during practice. If both operand orders of each division problem had been presented, each product would have occurred in two division problems and thus would have been associated with two different answers. The unique status of the product–answer associations would no longer have been present and neither the RT advantage for division by the end of practice, nor the division-specific interference effects at test discussed above, would be expected to occur.

*Test*

As predicted by the identical elements model, there were substantial performance differences between problems with noncritical and critical changes, but there was no evidence of differences among problems with critical changes for either multiplication or division. The apparent performance equivalence among new problems and problems in the other critical change conditions, in conjunction with the Campbell (1987) and Pauli et al. (in press) findings of no positive transfer to a new problem at test, is consistent with the identical elements assumption that independent knowledge units support retrieval performance on complementary multiplication and division problems, and also on the two "operand orders" of a division problem (e.g., 4 × _ = 28 and 7 × _ = 28).

## Experiment 2

One purpose of this experiment was to explore the extent to which the advantage of no change over operand order change problems for multiplication in Experiment 1 (see also Fendrich et al., 1993, and Rickard et al., 1994) reflects solely a perceptual-level processing advantage for the practiced operand order, as implied by the identical elements model. Participants were given practice on one operand order of a given problem in one format (e.g., six × nine) and were then tested on both operand orders in the practice format (e.g., six × nine and nine × six), as well as on both operand orders in a different format (e.g., 6 × 9 and 9 × 6), and on new problems not seen during practice in either operand order or either format (see Table 3). On the basis of the previous empirical findings, we expect an advantage for the practiced operand order over the unpracticed operand order when problems are presented in the practiced format. The identical elements model makes two additional predictions. First, there should be a substantial performance advantage for all noncritical change problems over critical changes problems (new problems), even when the format, operand order, or both are changed from practice to test. Second, any performance advantage for no change over operand order change problems should disappear when these problems are presented in an unpracticed format. This second prediction follows from the model because it predicts equivalent perceptual and cognitive stage processing for complementary operand orders of a practiced problem when that problem is presented in an unpracticed perceptual format. Consider for example the case in which the problem 4 × 7 is seen during practice. Performance on the reversed operand order problem in the same format (7 × 4) at test will be 50 to 100 ms slower on average than performance on the no change problem (i.e., the same problem seen during practice). The only way the identical elements model can fully account for this effect is to assume that the disadvantage for the operand order change problems reflects slower processing exclusively within the perceptual stage. Now consider the format change (four × seven) and format plus operand order change (seven × four) versions of the same problem at test. Because neither of these problems was seen in this format during practice, there is no reason to expect any differences between them in processing speed within the assumed perceptual stage. Thus, the identical elements model is constrained to predict no difference in performance on average across problems in the format and format plus operand order change conditions.

Table 3
*Test Conditions for an Example Problem at Test in Experiment 2*

| Test condition | Practice | Test |
|---|---|---|
| **Arabic format** | | |
| No change | 4 × 7 | 4 × 7 |
| Order change | 7 × 4 | 4 × 7 |
| Format change | four × seven | 4 × 7 |
| Order and format change | seven × four | 4 × 7 |
| New problems | | 4 × 7 |
| **Written verbal format** | | |
| No change | four × seven | four × seven |
| Order change | seven × four | four × seven |
| Format change | 4 × 7 | four × seven |
| Order and format change | 7 × 4 | four × seven |
| New problems | | four × seven |

This experiment also allows us to gain insight into the proportion of speed-up with practice that is specific and the proportion that is nonspecific. Specific speed-up reflects speed-up directly tied to the format in which problems are presented. Nonspecific speed-up refers to any remaining speed-up effect that cannot be directly associated with the format of presentation. We define specific speed-up operationally as the RT difference between the no change and format change conditions within a given format at test, and we define nonspecific speed-up as the difference between the format change and new problems conditions within each format. Under a strict additive factors formulation of the identical elements model, specific speed-up corresponds to speed-up within the perceptual stage, and nonspecific speed-up corresponds to speed-up within the cognitive stage.

The model makes three primary predictions about the pattern of specific and nonspecific speed-up. First, the majority of speed-up should be nonspecific, reflecting faster access to abstract problem presentations and the actual answer retrieval process. Second, because the empirical evidence indicates that the problem-size effect (for a review see Ashcraft, 1992) reflects differences in time required for answer retrieval, any reduction in the problem-size effect with practice should primarily be reflected as nonspecific speed-up. Third, although the model allows for the possibility of differences in specific speed-up for the two problems formats, there should be no difference in nonspecific speed-up for the two formats because nonspecific speed-up is assumed to reflect access to the unitary cognitive stage of processing for both formats.

One potential problem with the assumption that nonspecific speed-up in this experiment corresponds to cognitive stage speed-up is that the motor response sequences for problems in the format-change condition of the test will be familiar to participants (they will encounter them throughout practice), whereas the sequences for the new problems condition will be novel. Thus, any performance differences between these conditions might be related at least in part to differences in motor RTs, as well as to differences in the presumed cognitive stage of processing of primary interest. To control for this potential confound, we included digit entry trials throughout practice, on which participants simply entered the two-digit sequences corresponding to the answers to the new problems to be presented during the test.

## Method

### Participants

Twenty-four introductory psychology students participated in the experiment for course credit.

### Apparatus

Participants were tested on Zenith Data Systems personal computers, programmed with the MEL software (Schneider, 1988).

### Materials and Procedure

The materials and design of this experiment were identical to Experiment 1, with only the following changes. First, all problems were multiplication. The multiplication–division comparison of Experiment 1 was replaced by a comparison of problems in arabic versus written verbal format. Thus, there were the following five test conditions for each format: no change, operand order change, format change, format plus operand order change, and new problems (see Table 3). Second, each block consisted of two miniblocks. For 12 of the participants, the first miniblock throughout practice consisted of one exposure to each of the 16 multiplication problems in the participant's practice set and the second miniblock consisted of a set of 7 digit entry trials. The order of the miniblocks was reversed for the remaining 12 participants. The digit entry miniblocks equated the frequency with which each of the answers to be encountered at test were entered on the keypad during practice. Note that at test there was a total of 21 double-digit multiplication answers (three problem pairs shared the same answer). For each participant, 14 of these were entered on the key pad during practice in the context of the 16 multiplication problems. The remaining 7 answers to problems to be presented in the new problems were included in the digit entry miniblocks. The procedure for presenting the numbers in the digit entry miniblocks was analogous to that for presenting the multiplication problems: The two-digit numbers appeared on the screen one at a time, and the participants entered them on the number keypad. Before each miniblock, participants were prompted as to whether arithmetic or digit entry was required. Participants received two sessions of practice (30 blocks per session) on Monday and Wednesday of one week, followed by a test session on Friday of the same week. Finally, there were six blocks of test trials, where each of the 24 problems was presented once in each operand order and in each format in each block.

## Results

Unless otherwise noted, all data analysis procedures were as discussed in the first paragraph of the *Results* section of Experiment 1. Analyses done separately by problem difficulty yielded the same set of reliable and nonreliable results as did the overall analyses reported below.

### Practice

The overall error proportions for digit entry trials in Sessions 1 and 2 were .0278 and .0267 for the arabic format and .0257 and .0277 for the written verbal format, respectively. For multiplication problems, these values were .0378 and .0388 for arabic format and .0437 and .0437 for the written verbal format, respectively. Figure 4 shows log RTs for correctly solved problems plotted by log block, format (arabic or written verbal), and problem type (digit entry or multiplication). Each digit entry data point represents up to 96 observations, and each multiplication data point represents up to 192 observations.

Separate within-subjects regressions were performed on the digit entry and multiplication data with log block as a continuous variable and format (arabic or written verbal) as a categorical variable. For digit entry, there was a reliable overall effect of log block, $F(1, 23) = 152.3$, but no reliable interaction of log block and format, $F(1, 23) = 2.91$. There was a reliable performance advantage for arabic numbers both at the beginning of practice, $F(1, 23) = 6.97$, and at the end of practice, $F(1, 23) = 6.41$. For multiplication, there was a reliable overall effect of log block, $F(1, 23) = 206.8$, and a reliable interaction of log block and format, $F(1, 23) = 10.36$. There was a reliable advantage for arabic problems both at the beginning of

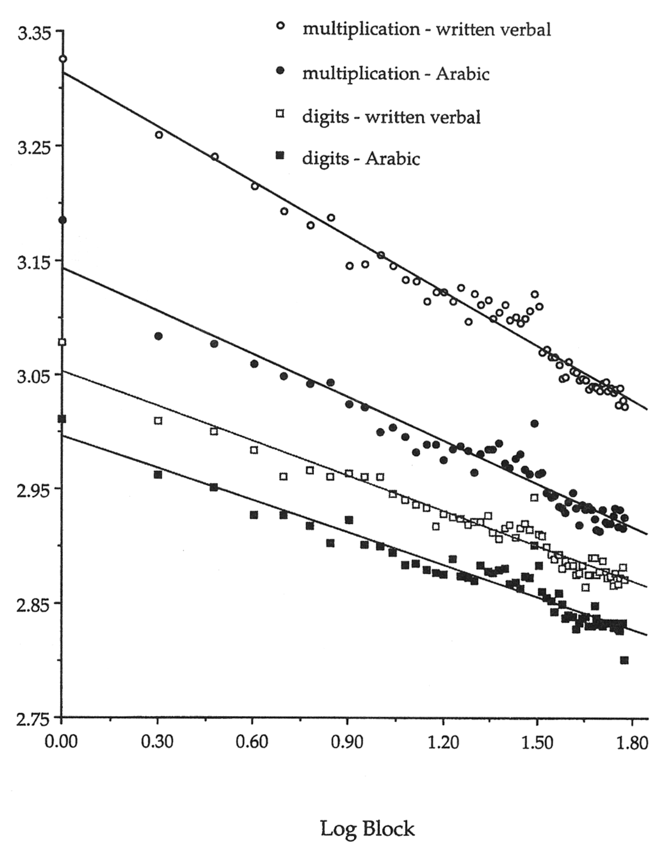


*Figure 4.* Log RT for correctly solved problems and digit entry trials in Experiment 2 plotted as a function of log block and format. RT = response time.

practice, $F(1, 23) = 21.26$, and at the end of practice, $F(1, 23) = 19.04$.

### *Test: Arabic*

The error proportions are plotted in Figure 5a by test condition. There was a marginal overall effect of test condition in a within-subjects ANOVA collapsed across block, $F(4, 23) = 2.32$, $p = .063$. Four planned orthogonal contrasts were performed. The first, comparing noncritical change conditions (i.e., the no change, operand order change, format change, and operand order plus format change conditions) with the critical change condition (i.e., the new problems condition), $F(1, 23) = 4.75$, and the second, testing for an overall effect of format (i.e., the average of the no change and operand order change condition compared with the average of the format and format plus operand order change conditions), $F(1, 23) = 4.1$, were both reliable. The third, comparing the no change and format change conditions with the operand order change and format plus operand order change conditions, $F(1, 23) = 0.14$, and the fourth, testing for an interaction of operand order (same or different) and format (same or different), $F(1, 23) = 0.28$, were not reliable. In summary, the only detectable differences were between noncritical and critical change conditions, and between same- and different-format conditions.

The anti-log of the mean log initiate RT is plotted in Figure 5b by block (1–6) and test condition. A 6 × 5 ANOVA with within-subjects variables of block and test condition revealed reliable effects of test condition, $F(4, 23) = 27.28$; block, $F(5, 23) = 22.46$; and a reliable interaction of test condition and block, $F(20, 23) = 4.78$. The interaction reflects generally increasing rates of speed-up during practice with increasing overall difficulty of the test condition. Four planned orthogonal contrasts (identical to those performed on the error data) were performed on the test condition variable. The first,

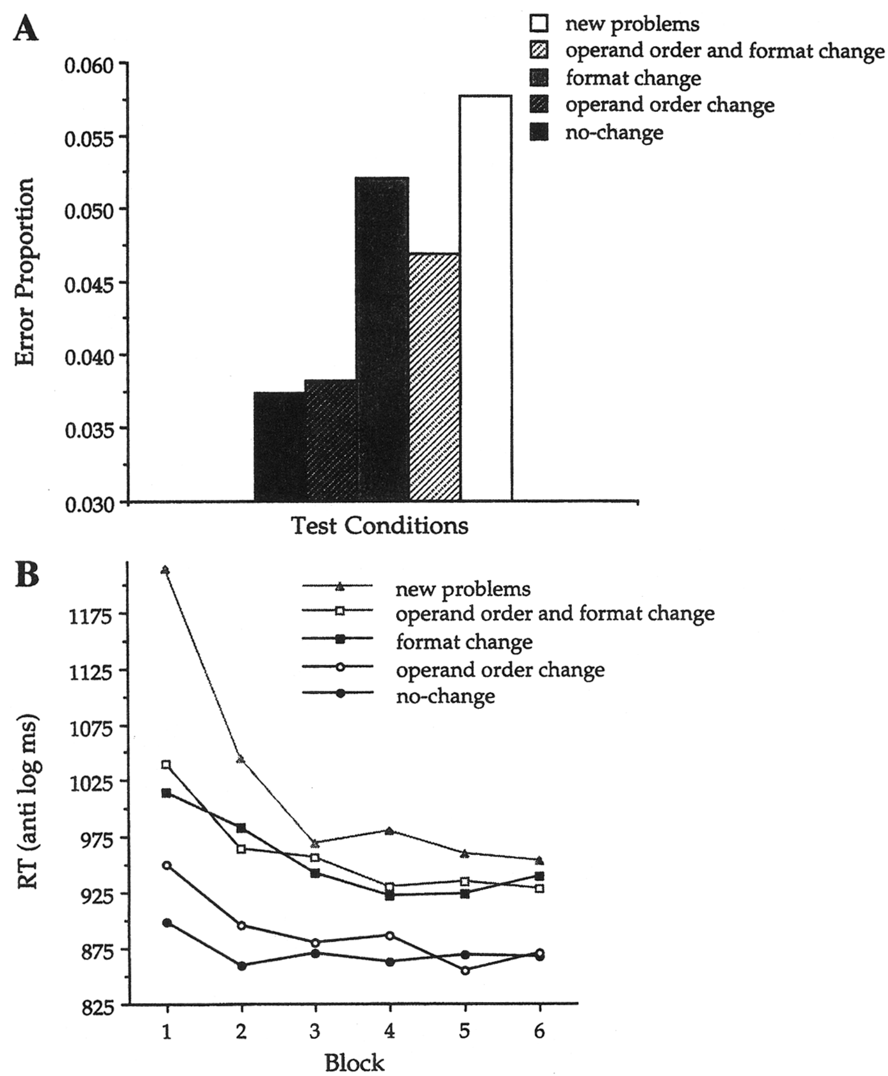


*Figure 5.* A: Error rates and B: anti-log RTs for arabic problems in Experiment 2. Error data are collapsed across test blocks. RT = response time.

comparing noncritical change conditions with the critical change condition, was strongly reliable, $F(1, 23) = 59.34$. An additional post hoc contrast comparing the format and format plus order change conditions with the new problems condition was also reliable, $F(1, 23) = 17.63$, providing further verification of the advantage for noncritical compared with critical change conditions. The second planned contrast, testing for an overall effect of format (same or different), was also reliable, $F(1, 23) = 48.17$. The third, testing for an overall effect of operand order (same or different), $F(1, 23) = 1.12$, and the fourth, testing for an interaction of operand order and format, $F(1, 23) = 0.5$, were not reliable. The failure to find either a main effect of operand order or an interaction between operand order and format suggests that there were no operand order effects in the RT data. More focused post hoc analyses, however, comparing the no change and operand order change conditions separately for each block of test, showed reliable operand order effects for Block 1, $F(1, 23) = 12.81, p = .002$, and Block 2, $F(1, 23) = 7.62, p = .01$, but no reliable effects for the remaining blocks. Thus, there does appear to have been an RT advantage for no change over operand order change problems, although, for reasons unknown, the effect was less persistent than in previous experiments (Experiment 1 of this study; Fendrich et al., 1993; Rickard et al., 1994).

### *Test: Written Verbal*

The error proportions are plotted in Figure 6a by test condition. An ANOVA with a five-level within-subjects variable of test condition revealed a reliable overall effect of test condition, $F(4, 23) = 5.84$. Four planned orthogonal contrasts, identical to those performed for the arabic format, were performed on the test conditions. The first, comparing noncritical change conditions with the critical change condition, was reliable, $F(1, 23) = 14.78$. The second, testing for an overall effect of format, $F(1, 23) = 0.42$, and the third, testing for an overall effect of operand order, $F(1, 23) = 0.86$, were not reliable. The fourth, testing for an interaction of operand order and format, was reliable, $F(1, 23) = 7.07$. The most reasonable interpretation of this interaction is that it reflects higher errors rates in the operand order, format, and format plus operand order change conditions, relative to the no change condition.

The anti-log of the mean log initiate RT is plotted in Figure 6b by block (1–6) and test condition. A 6 × 5 ANOVA with within-subjects variables of block and test condition revealed reliable effects of test condition, $F(4, 23) = 48.39$, and block, $F(5, 23) = 17.72$, and a reliable interaction of test condition and block, $F(20, 23) = 3.13$. Four orthogonal contrasts (identical to those performed for the error analysis) were performed on the test condition variable. The first, comparing noncritical change conditions with the critical change condition, was strongly reliable, $F(1, 23) = 77.08$. An additional post hoc contrast comparing the format and format plus operand order change conditions with the new problems condition was also reliable, $F(1, 23) = 17.42$. The second contrast, testing for an overall effect of format, $F(1, 23) = 88.54$, and the third, testing for an overall effect of operand order, $F(1, 23) = 18.78$, were both reliable, as was the interaction contrast, $F(1, 23) =$ 9.17. Additional nonorthogonal contrasts indicated that the effect of operand order was reliable given no change in format, $F(1, 23) = 27.1$, but was not reliable given a change in format, $F(1, 23) = 0.85$.

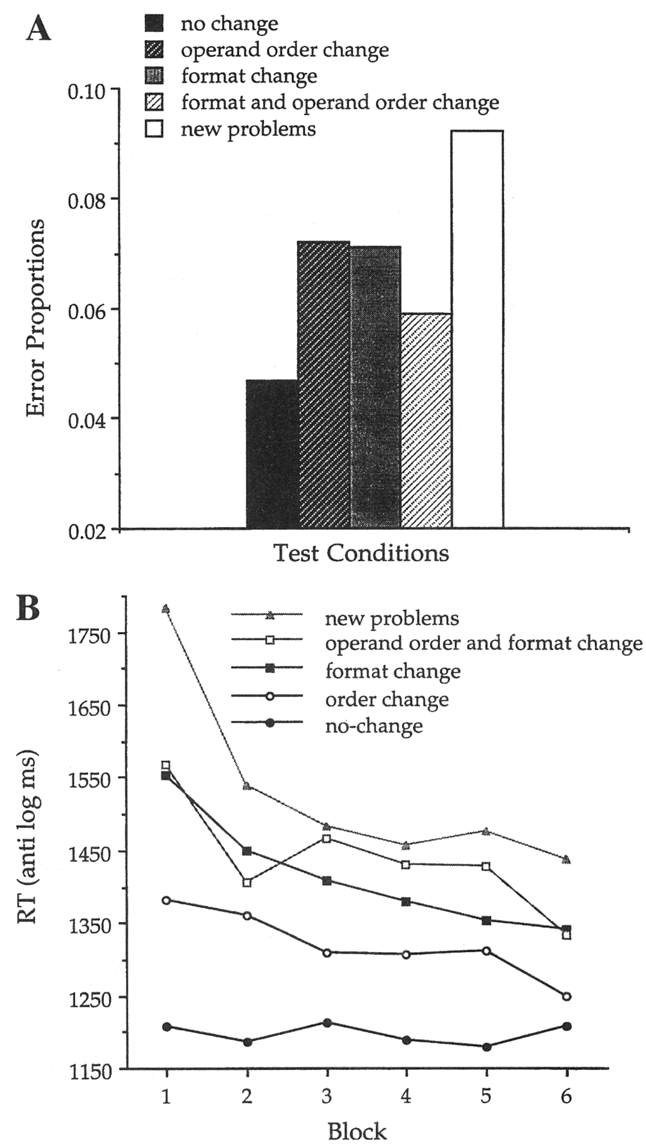


*Figure 6.* A: Error rates and B: anti-log RTs for written verbal problems in Experiment 2. Error data are collapsed across test blocks.

One other test result worth noting is that the effect of a simple operand order reversal depended strongly on format. There were only negligible effects of reversing operand order for arabic problems, but quite substantial effects for written verbal problems. One potential account of this difference is that arabic problems are parsed into their constituent visual components (e.g., 4, ×, and 7), and that processing of these visual components takes place largely in parallel. Thus, no operand order effects would be observed for arabic problems. The system that decodes written verbal problems, however, might process the problems elements sequentially. Written verbal problems subtended a larger visual angle in this experiment, making a sequential decoding process quite plausible. If these representations, which may take a phonological form, are distinct for each operand order, then specific speed-up with practice on one operand order would not be expected to transfer significantly to the complementary order. Note that these hypotheses do not account for all of the observed operand order effects, as there was a small difference between no change and operand order change problems in the arabic format, and also a slight advantage for operand order change problems over format change problems in the written verbal format. However, it is clear at least that representations that mediate performance on written verbal problems are much more sensitive to operand order than are those that mediate performance on arabic problems.

### *Test: Specific Versus Nonspecific Speed-Up*

A separate analysis of the tests data was performed to investigate perceptually specific and nonspecific components of improvements in RT with practice. This analysis was limited to the first block of test to avoid contamination from possible differences among the test conditions in rate of RT improvement over test blocks. Raw RTs were used as the dependent variable to preserve the validity of RT predictions of an additive factors interpretation of the identical elements model. As discussed previously, the difference between format change and no change conditions can be taken to reflect the specific component of speed-up due to practice. Because there were no reliable differences between the format and format plus operand order change conditions in the overall test results, these conditions were collapsed together for this analysis. Similarly, the difference between new problems and the average of format change and format plus operand order change problems can be taken to reflect the nonspecific component of speed-up.

Two other factors were considered in the analysis. First, a problem-size factor was created by dividing problems into easy and difficult categories on the basis of normative data from Campbell and Graham (1985). This factor allowed us to investigate two issues. First, is there a reduction in the magnitude of the problem-size effect with practice (i.e., will the problem-size effect be smaller for no change problems than for new problems)? Second, will any reduction in the problem-size effect with practice be observed only for specific or nonspecific components of speed-up, or for both of these components? The most straightforward prediction of the identical elements approach is that a reduction in problem-size effect will be observed only with respect to nonspecific component of speed-up. The third factor in this analysis was problem format (arabic or written verbal). This factor allowed us to explore whether there are any RT differences overall between the arabic and written verbal formats, and of more interest, whether any differences between format change and new problems conditions depend on format. An additive factors version of the identical elements model predicts no difference between formats for comparisons limited to these two test conditions. That is, the model predicts no difference in degree of nonspecific speed-up as a function of problem format, although it does allow for format-related differences in specific speed-up

Mean RTs averaged over participants are plotted in Figure 7 by test condition (no change, the average of format and format plus operand order change, and new problems), by format (arabic or written verbal), and by problem difficulty (easy or difficult). A 3 (test condition) × 2 (format) × 2 (problem difficulty) within-subjects ANOVA revealed reliable effects of test condition, $F(2, 23) = 66.4$; format, $F(1, 23) = 98.7$; and problem difficulty, $F(1, 23) = 31.9$. There was also a reliable two-way interaction between test condition and problem size, $F(1, 23) = 6.19$, confirming an overall decrease in the problem-size effect for no change problems compared with new problems. No other interactions reached significance (all $F$s $< 1$).

Two more focused analyses were performed to further explore specific and nonspecific components of speed-up with practice. First, to explore specific speed-up effects, the ANOVA described above was performed limited to the no change and format change test conditions. The main effects of test condition, $F(1, 23) = 79.3$; format, $F(1, 23) = 125$; and problem size, $F(1, 23) = 47$, were all reliable. The main effect of test condition confirms an overall specific speed-up effect. Of more interest, the interaction between test condition and format was significant, $F(1, 23) = 7.57$, indicating more specific speed-up with practice for problems presented in the written verbal format. There was also a nonsignificant trend toward a Test Condition × Problem-Size interaction, $F(1, 23) = 3.57$, $p =$

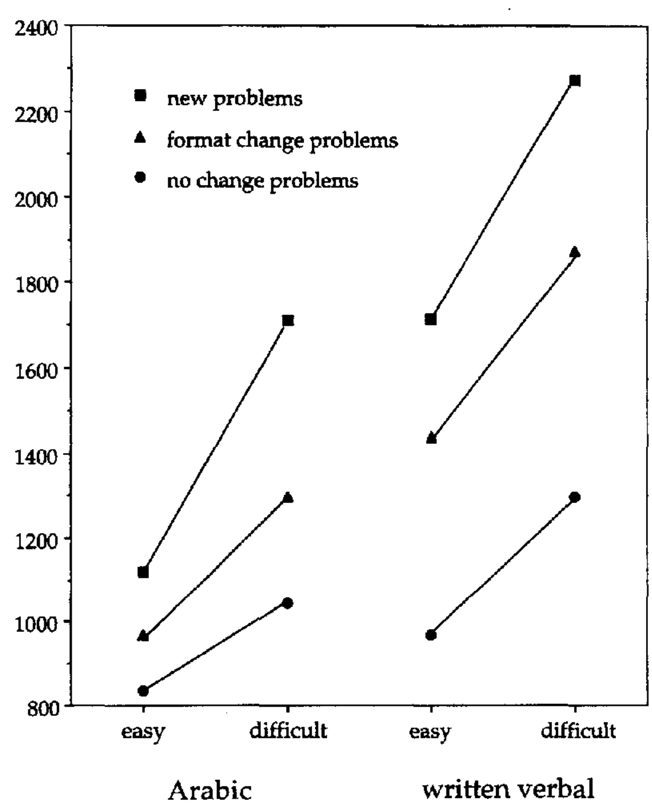


*Figure 7.* Raw RTs on the first block of test in Experiment 2 as a function of format, test condition, and problem difficulty.

.071, suggesting that some of the reduction in the problem-size effect with practice reflects the specific component of processing, although follow-up work would be needed to confirm this trend. None of the remaining interactions were reliable (all $Fs < 1$).

An additional ANOVA was performed on the format change and new problems test conditions. The main effect of test condition, $F(1, 23) = 44.5$; format, $F(1, 23) = 72.4$; and problem size, $F(1, 23) = 23.6$, were all reliable. The main effect of test condition confirms an overall nonspecific speed-up effect. The interaction between test condition and problem size was also reliable, $F(1, 23) = 6.2, p = .021$, indicating that at least part of the reduction of the problem-size effect with practice reflects the nonspecific component of processing. In contrast, the Test Condition × Format interaction, which was reliable in the preceding analysis, was not reliable in this analysis, $F(1, 23) = 0.18$. Thus, improvement in performance reflected in the nonspecific component of processing does not appear to depend on format. Neither of the remaining interactions reached significance ($Fs < 1$).

These results are generally consistent with the predictions of the identical elements model. First, there was substantial nonspecific speed-up that actually exceeded the specific speed-up for arabic problems. Second, there was a reduction in the magnitude of the problem-size effect evident in the nonspecific component of speed-up. Third, although the degree of specific speed-up clearly depended on format, the nonspecific speed-up apparently did not.

## General Discussion

From the theoretical perspective of the identical elements model, all of the salient test results from both experiments can be summarized simply: (a) There was substantially better performance in noncritical change than in critical change test conditions, (b) there were no reliable performance differences among critical change conditions, and (c) when differences in perceptual familiarity were controlled for, there were no reliable performance differences among noncritical change conditions. Moreover, the specific versus nonspecific speed-up results are consistent with the identical elements prediction that there are distinct specific (perceptual) and nonspecific (cognitive) components of processing, that the problem-size effect is reliably reflected in the nonspecific component, and that format of presentation does not influence the nonspecific component.

There are, however, two findings from the component speed-up analysis that are potentially problematic for the model. First, there was in fact a quite large specific component of speed-up. A more natural prediction of the model is that a majority of speed-up should be within the nonspecific component. Another potential problem is the strong trend ($p = .07$) indicating a reduction in the problem-size effect with practice within the specific component. Both of these results suggest that there is more going on within the specific component of processing than would be expected a priori within the framework of our model. Conceivably, however, an assumption that the perceptual stage processing speeds up considerably as a function of the frequency with which a problem is encountered, combined with an assumption that small problems are seen more frequently than large problems, can account for both the substantial specific speed-up and the trend toward more specific speed-up for large problems. Alternatively, the identical elements model may require elaboration. One possibility is that an abstract representation mediates number-fact retrieval in most everyday circumstances, but that more direct connections bypassing this stage may develop under conditions of extensive practice on a small set of problems that are always encountered in the same input format and with the same output requirements. This dual-route approach (with obvious similarities to well-known reading models) lacks parsimony, but it does account naturally both for our findings that suggest access to an abstract representation, as well as for those that suggest more direct, or asemantic, connections.

There have been two recent patient studies that are also relevant to evaluating the identical elements model. Hittmair-Delazar, Semenza, and Denes (1994) discussed a patient who exhibited a relatively severe arithmetic fact retrieval deficit. Consistent with the identical elements model, there was essentially no difference for complementary operand orders in terms of whether fact retrieval or a calculational strategy was reported by this patient. In other words, the fact retrieval strategy was available for one operand order if and only if it was also available for the other operand order. McCloskey, Aliminosa, and Sokol (1991) found similar effects in terms of error patterns.

Hittmair-Delazar et al. (1994) also found that training of this patient on one operand order of each problem transferred only slightly to the reversed-order problems, suggesting that speed-up observed with practice reflected strengthening of operand-order-specific problem representations. This finding is not necessarily inconsistent with the identical elements model, however, for two reasons. First, RTs were very long for this study, in the range of 3.5 to around 8.0 s. This range is well outside of that of our experiments and suggests that calculation strategies, rather than fact retrieval, may have predominated. Second, the abstraction processes that we assume generalizes across operand order may only be able to operate when both operand orders of a problem are encountered repeatedly over the learning interval during which the transition from use of calculational strategies to fact retrieval occurs. Thus, if calculation strategies were used by Hittmair-Delazar et al.'s patient at the beginning of practice, any transition to retrieval that may have occurred during practice might reflect development of operand-order-specific memory structures. The identical elements model only predicts strong transfer across operand order if performance on both operand orders reflects fact retrieval at the outset of the practice interval.

The reasonable success of the identical elements model in accounting for the existing data invites comparison with alternative frameworks. One broad theory of practice effects that has been tested using arithmetic tasks is the instance theory of automaticity (Logan 1988). Our results would appear to be relevant to determining exactly what constitutes an instance in memory in this task domain. Complementary problems from different operations clearly correspond to a separate set of instances. However, whereas the data from Siegler (1986) indicate that children often represent multiplication problems initially in an operand-order-specific way, the data from this study show that these representations become generalized across operand order after sufficient practice. Such transformations are not predicted by the current version of the instance theory, and they suggest that the instance approach to memory will need to be modified or extended to account for the entire range of changes in memory and skilled performance that occur with practice in this domain.

The encoding-complex approach to number processing proposed by Campbell and Clark (1992) assumes that problems are represented by specific physical codes tied to the perceptual modality through which the problems are processed. There is no abstract or generalized form of representation in this framework. Rather, the various physical codes are directly connected, and each can in principle participate in fact retrieval regardless of the format or modality in which a problem is presented. This framework can account for the strong specific speed-up observed in both formats and also for the trend suggesting a reduction in the problem-size effect within the specific component of processing. The framework also accounts for the nonspecific speed-up under the assumption that specific codes other than those directly corresponding to the format of presentation are participating in answer retrieval. However, our finding of more nonspecific than specific speed-up for problems presented in the arabic format is not directly predicted by the encoding-complex approach. This finding implies that, at least under some circumstances, speed-up with practice reflects largely or even primarily the influence of specific codes other than those directly corresponding to the format of presentation. Note that this finding holds under the conditions of Experiment 2, in which general transfer factors such as changes in global speed–accuracy criteria and response production processes are controlled. Thus, this result places potentially useful empirical constraints on the dynamics of the interaction among the various specific codes in the encoding-complex approach.

Another recent proposal is Deheane's (1992) triple code theory, which among other things proposes a verbal word code into which a problem presented in any format must be transcoded prior to answer retrieval. Predictions of this model are similar to those of the identical elements model with respect to the current experiments. Note, however, that it is unclear how a verbal word code would account for our finding of no effect of the operand order factor given a change in format. If this code is related at all to a phonological representation, then it is likely to have a strong sequential character and thus complementary operand orders would map onto different codes. Thus, the most obvious prediction of a verbal word code is that there should be an RT advantage for the practiced over the unpracticed operand order even with a change in format at test, an effect that was not observed.

The reasonable success of the identical elements model for basic multiplication and division also raises questions about its generalizability to other tasks. One obvious question is whether the identical elements model will work for basic addition and subtraction. Predictions of the model for addition match exactly with those for multiplication, and those for subtraction match exactly with those for division. It is unclear a priori whether these predictions will prove accurate, as there is some evidence suggesting that skilled addition and subtraction are mediated by an entirely different form of representation. For example, people may rely more on mental manipulation of visuospatial number line segments when performing these operations (see Koshmider & Ashcraft, 1991; Restle, 1970). Transfer experiments similar to those described in this article, but which explore addition and subtraction, might prove useful in discriminating between a discrete facts model like the identical elements model and models emphasizing analog processes, such as that of Restle (1970).

In principle, the predictions of the identical elements model should also hold for the commonly employed multiplication verification task (i.e., $4 \times 7 = 35$; True of False?). Improvements with practice should transfer strongly when operand order is reversed. However, there should be no transfer across operation. For example, speed-up with practice verifying that 28 is a true answer for the problem $4 \times 7$ should not improve performance verifying that 7 is a true answer for $28 \div 4$. Note, however, that there appears to be a wide variety of strategies available for verification that are not viable for production (Romero, Rickard, & Bourne, 1994; Zbrodoff & Logan, 1986), and some of these appear to allow the participant to circumvent answer retrieval altogether. If efficiency or relative frequency of use of generic nonretrieval strategies changes with practice, it may prove difficult to test the identical elements model using a verification transfer experiment.

Another central issue is whether the model can account for changes due to practice when initial skill levels are substantially lower than in the current experiments. Rickard and Bourne (1995) addressed this question by exploring the effects of practice on a totally novel math task built on the basic arithmetic operations. The task used in this experiment required using a generic calculation algorithm initially, but participants exhibited a transition to direct retrieval of answers from memory after extended practice. They were then tested on problems that corresponded to problems in the no change, operation change, and new problems conditions in the current experiments. As predicted by the identical elements model, the memory retrieval strategy at test was only used on no change problems, and RTs and error rates for operation change and new problems were several times greater than those for no change problems. These results demonstrate clearly that the identical elements prediction of no transfer between complementary operations holds even for novel tasks.

Although the identical elements model is intended to account only for long-term memory organization, it is interesting to speculate on whether the predictions of the model would hold under conditions that might support automatic short-term priming. For example, would answering $28 \div 4 = \_$ facilitate RT for $4 \times 7 = \_$ on the immediately following trial? In this example the perceptual overlap between the problems is minimal, so it is conceivable that, at least under conditions in which the participant does not anticipate that the preceding problem can provide a cue, no such priming would be observed. This type of experiment might prove useful for exploring the degree of specificity of automatic facilitation priming involved in fact retrieval.

As one final note it is worth pointing out that our model does not make predictions about interactions among problem representations that might produce interference, and thus it is not inconsistent with established effects like negative transfer to new problems at test (Campbell, 1987) or interproblem interference effects observed in Experiment 1 of this study, as well as in many other studies (e.g., Campbell & Graham, 1985). The model is also silent on the details of any associative network or connectionist system that might implement the assumed abstract level of representation. It is at this level of detail that both the predictions of the identical elements model and the error and interference patterns referred to above, could potentially be integrated. The focus of the model at present is the conditions under which positive transfer, indexed either by RTs or error rates, should or should not be observed. Thus, even if the model proves to be correct, clearly it addresses only a piece of the broader puzzle of how basic arithmetic is represented and processed.

## Conclusions

The test data from both experiments support the following new empirical conclusions regarding adult arithmetic skills: (a) For educated adults, improvements in performance with practice do not transfer from multiplication to division, or vice verse; (b) improvements do not transfer between complementary operand order problems within division (e.g., practice on 4 × _ = 28 does not transfer to 7 × _ = 28); (c) these improvements reflect both perceptually specific and perceptually nonspecific processing components; (d) the problem-size effect is clearly reflected in the nonspecific component of processing, and perhaps in the specific component as well; (e) differences in speed-up with practice tied to format of presentation appear to be confined to the specific component of processing; (f) for multiplication there is more transfer to reversed order problems for the arabic format than for the written verbal format; and (g) practice transfers completely to reversed order problems once perceptually specific practice effects are controlled. In summary, these results generally support an identical elements model of the organization of arithmetic skill in memory. This model should provide a useful reference point for future research because it makes relatively simple and concise empirical predictions that could potentially be descriptive of a variety of tasks.

## References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451–474.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44,* 75–106.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 109–123.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53,* 1–44.

Campbell, J. I. D., & Clark, J. M. (1992). Cognitive number processing: An encoding-complex perspective. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 457–491). Amsterdam: Elsevier.

Campbell, J. I. D., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology, 39,* 338–366.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

Fendrich, D. W., Healy, A. F., & Bourne, L. E., Jr. (1993). Mental arithmetic: Training and retention of multiplication skill. In C. Izawa (Ed.), *Applied cognitive psychology: Applications of cognitive theories and concepts* (pp. 113–133). Hillsdale, NJ: Erlbaum.

Hittmair-Delazar, M., Semenza, C., & Denes, G. (1994). Concepts and facts in calculation. *Brain, 117,* 715–728.

Koshmider, J. W., & Ashcraft, M. H. (1991). The development of children's multiplication skill. *Journal of Experimental Child Psychology, 40,* 49–72.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

McCloskey, M. (1992). Cognitive mechanisms in number processing: Evidence from acquired dyscalculia. *Cognition, 44,* 107–157.

McCloskey, M., Aliminosa, D., & Sokol, S. M. (1991). Facts, rules, and procedures in normal calculation: Evidence from multiple single-case studies of impaired arithmetic fact retrieval. *Brain and Cognition, 17,* 154–203.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4,* 171–196.

Pauli, P., Lutzenberger, W., Birbaumer, N., Rickard, T. C., & Bourne, L. E., Jr. (in press). Neuropsychological correlates of mental arithmetic. *Psychophysiology.*

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83,* 274–278.

Rickard, T. C., & Bourne, L. E., Jr. (1995). An identical elements model of basic arithmetic skills. In A. F. Healy, & L.E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills* (pp. 255–281). Newbury Park, CA: Sage.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1139–1153.

Romero, S., Rickard, T., & Bourne, L. E., Jr. (1994, November). *On verification of multiplication facts: An investigation using retrospective protocols.* Poster session presented at the 35th annual meeting of the Psychonomic Society, St. Louis, MO.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, and Computers, 20,* 206–217.

Siegler, R. S. (1986). *Children's thinking.* Englewood Cliffs, NJ: Prentice-Hall

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General, 115,* 118–130

Received June 13, 1995
Revision received August 7, 1995
Accepted January 16, 1996 ■