Memory & Cognition
2006, 34 (1), 217-226

Calculation latency: The μ of memory and the τ of transformation

JAMIE I. D. CAMPBELL
University of Saskatchewan, Saskatoon, Saskatchewan, Canada

and

MARCIE PENNER-WILGER
Carleton University, Ottawa, Ontario, Canada

Does numeral format (e.g., 4 + 8 vs. four + eight) affect calculation per se? University students ($N = 47$) solved single-digit addition problems presented as Arabic digits or English words and reported their strategies (memory retrieval or procedures such as counting or transformation). Decomposition of the response time (RT) distributions into μ (reflecting shift) and τ (reflecting skew) confirmed that retrieval trials contributed predominantly to μ, whereas procedure trials contributed predominantly to τ. The format × problem size RT interaction (i.e., greater word-format RT costs for large problems than for small problems) was associated entirely with μ and not with τ. Reported use of procedures presented a corresponding format × size interaction. Together, these results indicate that, relative to the well-practiced digit format, the unfamiliar word format disrupts number-fact retrieval and promotes use of procedural strategies.

The effect of numeral format (e.g., Arabic digits, written number words, Roman numerals) on performance of simple arithmetic continues to be a controversial issue in numerical cognition research. One view is that effects of format are restricted to systems that encode numerical stimuli, and do not penetrate downstream to affect calculation (Dehaene & Cohen, 1995; McCloskey, 1992; Noël, Fias, & Brysbaert, 1997). Others have reported evidence that format can directly affect the efficiency of calculation per se (Blankenberger & Vorberg, 1997; Campbell, 1994; Campbell, Kanz, & Xue, 1999; Campbell, Parker, & Doetzel, 2004; McNeil & Warrington, 1994; Sciama, Semenza, & Butterworth, 1999). Here, we pursue this issue by analyzing parameters of the latency distributions produced by digit and word problems. Specifically, we estimated μ, which is sensitive to a uniform shift in the entire response time (RT) distribution, and τ, which is sensitive to shifts in skew (Smith & Mewhort, 1998). We will begin by briefly reviewing a key phenomenon in the debate, and then explain our predictions regarding effects of format on RT distributions.

Format × Problem Size Interaction in Cognitive Arithmetic

Numerous experimental studies of simple addition and multiplication have compared performance with problem operands presented as Arabic digits to performance with operands as written number words. A wide range of languages have been examined, including French, Dutch, English, German, Chinese, and Filipino (see Campbell & Epp, 2005, for a review). One general finding is that performance of simple arithmetic with written number words (e.g., seven × six) is much more difficult than with Arabic numerals (e.g., 7 × 6). Simple arithmetic with written words is as much as 30% slower and 30% more error prone than with Arabic digits (Campbell, 1994). Thus, arithmetic performance with the less-typical written word format is substantially impaired relative to the typical Arabic format.

Evidence that format affects calculation per se, rather than only more peripheral encoding or response processes, comes from the *problem-size effect* (PSE). The PSE is the common finding that the difficulty of simple arithmetic problems generally increases with the numerical size of the operands. Small-number problems have greater memory strength because they are encountered more frequently and may be less susceptible to retrieval interference (see Zbrodoff & Logan, 2005, for a detailed discussion of the PSE). Several studies have demonstrated, however, that the PSE in both simple addition and multiplication is larger with problems in written word format (e.g., three + eight) than in digit format (3 + 8; Campbell, 1994; Campbell & Clark, 1992; Campbell et al., 1999; Noël et al., 1997). For example, Campbell (1994) contrasted performance on *small* addition problems (both

---

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. We thank Steven Hecht, Patrick Lemaire, and an anonymous reviewer for very helpful feedback on a previous version of the manuscript. We also thank Craig Leth-Steensen for guidance using RTSYS. Address correspondence to J. I. D. Campbell, Department of Psychology, 9 Campus Drive, University of Saskatchewan, Saskatoon, SK, S7N 5A5 Canada (e-mail: jamie.campbell@usask.ca) or to M. Penner-Wilger, Institute of Cognitive Science, 1125 Colonel By Drive, Carleton University, Ottawa, ON, K1S 5B6 Canada (e-mail: wilger@sympatico.ca).

operands ≤5) to performance on *large* addition problems (at least one operand >5), and found that the RT PSE was much larger with English number words (192 msec) than with Arabic digits (108 msec). Given that the PSE arises during calculation, the format × size interaction suggests that format affects the efficiency of calculation, with greater word-format costs for more difficult problems. Retrieval processes for larger problems, which often have relatively weak problem–answer associations, might be especially susceptible to disruption by the unfamiliar-looking word format (Campbell, 1994; Campbell et al., 2004). Some researchers have argued, however, that the format × size interaction could arise during problem encoding if it takes an especially long time to encode larger number words for the purpose of calculation (McCloskey, Macaruso, & Whetstone, 1992; Noël et al., 1997; but see Campbell, 1999).

Campbell and Fugelsang (2001) demonstrated effects of format on calculation performance, however, which cannot be attributed to problem-encoding processes. In their experiment, adults solved simple addition problems in a true/false verification task with equations in digit format (e.g., 3 + 4 = 8) or written English format (three + four = eight). The format × size effect (i.e., greater word-format RT costs for large problems than for small problems) was nearly twice as large for true equations than for false equations. Encoding and response requirements were the same for true and false equations; consequently, the triple interaction cannot be attributed to effects arising at these stages. True verification trials are more sensitive to the efficiency of arithmetic processes than false trials, because arithmetic processing is less likely to need to run to completion for false equations (e.g., people often can reject a false equation without having to retrieve or calculate the correct answer; Lemaire & Reder, 1999); consequently, the larger format × size effect for true equations than for false equations suggests that the word format reduced the efficiency of arithmetic retrieval processes.

Participants' strategy reports in the Campbell and Fugelsang (2001) study reinforced this conclusion. After each trial, participants reported their solution strategy (direct retrieval or a procedural strategy such as counting or transformation; e.g., 6 + 7 = ? 6 + 6 = 12 + 1 = 13). Direct memory retrieval is the predominant strategy reported by educated adults for simple arithmetic (e.g., 6 + 3 = 9, 7 × 9 = 63), but many people report sometimes using procedural strategies, especially for larger simple arithmetic problems (Campbell & Xue, 2001; Geary & Wiley, 1991; LeFevre, Sadesky, & Bisanz, 1996). The effect of format on strategy reports was striking: Campbell and Fugelsang found that reported use of procedures was much greater with word stimuli (41% of all problems) than with digit stimuli (26%), and the increase in procedure use with words, when compared with digits, was greater for numerically larger problems. Because procedural strategies were about 500 msec slower, on average, when compared with retrieval, the disproportionate use of procedures for large, word-format problems contributed to the format × size interaction on RTs.

Campbell et al. (2004) extended this finding to the arithmetic production task (e.g., 7 + 8 = ?) by demonstrating that Canadian university students were 50% more likely to report procedural strategies (e.g., counting, transformation) for simple addition with written English operands when compared with Arabic digit operands. Unlike in Campbell and Fugelsang (2001), however, there was no evidence of a format × size interaction on use of procedures. Nonetheless, the main effect of format on procedure usage contributed directly to the format × size interaction in RT. This occurred because the PSE was substantially larger for procedure trials than for retrieval trials (Campbell & Xue, 2001; LeFevre, Sadesky, & Bisanz, 1996): Procedural strategies generally are slower than retrieval, but especially so for larger problems, which require more steps or more difficult steps than do smaller problems (LeFevre, Sadesky, & Bisanz, 1996). This inflates the PSE for procedures, when compared with retrieval. Consequently, the higher proportion of procedure trials with words, when compared with digits, would contribute to a larger PSE for words than for digits.

The strategy effects observed by Campbell and Fugelsang (2001) and Campbell et al. (2004) suggest that cognitive arithmetic is directly affected by surface form, because the unfamiliar word-problem format greatly increased reported use of procedural strategies. According to Siegler and Shipley's (1995) Adaptive Strategy Choice Model, selection of an arithmetic strategy depends on its relative efficiency (i.e., speed and probability of success). Consequently, any factor that reduces retrieval efficiency promotes a switch to procedures (Campbell & Timm, 2000). Less-efficient retrieval with the word format would therefore be expected to promote increased use of procedural strategies.

Format, Arithmetic Strategy, and RT Distributions

Campbell et al. (2004) proposed that the format × size RT interaction in simple arithmetic has two primary sources, both linked to the disruptive effects of the word format on memory retrieval. First, when participants use retrieval, it is slower for words than for digits, but especially so with the larger, more difficult problems. Second, participants are more likely to use nonretrieval procedures with the word format, and because procedural strategies produce a larger PSE than does direct retrieval, increased use of procedures for word problems contributes to a larger PSE for words than for digits (i.e., the format × size RT interaction).

The validity of the evidence supporting this view, however, hinges on the validity of participants' self-reported strategies. Self-reported estimates of retrieval versus procedure usage account consistently for RT differences across arithmetic operations, problem characteristics, and cultural groups (Campbell & Gunter, 2002; Campbell & Xue, 2001; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996), as well as across a variety of experimental manipulations, including effects of RT deadlines (Campbell & Austin, 2002), retrieval interference (Campbell & Timm, 2000), and manipulation of surface format (Campbell & Fugelsang, 2001; Campbell et al., 2004). Nonetheless, although the effects of format on strategy reports for simple arithmetic are striking, it is important to continue to evaluate the veridicality of the reports (Kirk & Ashcraft, 2001).

To pursue the validity of the strategy reports and the effects of surface format on number fact retrieval, we evaluated predictions about the shapes of the RT distributions produced by digit and word problems. Analysis of the RT distributions was done by applying the ex-Gaussian distributional model (Hockley, 1984; Hohle, 1965; Ratcliff & Murdock, 1976) using the software program RTSYS (Heathcote, 1996). The model provides a good fit to positively skewed RT data, including arithmetic data (for a review, see Penner-Wilger & Leth-Steensen, 2005). The ex-Gaussian model assumes a normally distributed component (corresponding to the main body of the RT distribution) and an exponentially distributed upper tail (corresponding to relatively slow RTs contributing to skew). The ex-Gaussian model provides quantitative estimates of μ, the mean of the normal component, and of τ, the mean of the exponential tail. Mean RT is simply the sum of these two components (mean RT = $\mu + \tau$). μ is sensitive to a uniform shift in the entire distribution; thus, an increase in μ is reflective of an overall slowing of responses. τ is sensitive to a shift in skew; thus, an increase in τ is reflective of a slowing associated with the upper tail of the distribution. Consequently, ex-Gaussian analysis can be used to test whether experimental manipulations selectively affect μ (i.e., shift the entire distribution) or τ (i.e., shift the upper tail).

Penner-Wilger, Leth-Steensen, & LeFevre (2002) used the ex-Gaussian model to examine differences in RT distributions across cultures, focusing on the PSE. Young adults educated in Canada or China solved single-digit multiplication problems. Campbell and Xue (2001) had already reported that the two groups differed in terms of the mix of solution strategies used: China-educated adults reported exclusive reliance on retrieval, whereas Canada-educated adults reported a mix of solution strategies. Penner-Wilger et al. found that the PSE for the Chinese group was reflected only in μ and not in τ, whereas the PSE for the Canadian group was reflected in both μ and τ. Changes in μ were posited to reflect changes in retrieval efficiency, and changes in τ were posited to reflect changes in the amount of procedure use or in the efficiency of procedure use.

The Present Experiment

In this experiment, young adults solved simple addition problems presented in either digit or word format and reported their strategy after each trial by selecting from remember, count, transform, or other. Theoretically, selection of the remember category corresponds to direct memory retrieval, whereas selection of one of the other categories corresponds to procedure use (cf. Campbell & Austin, 2002; Campbell & Fugelsang, 2001; Campbell et al., 2004; Campbell & Timm, 2000; Campbell & Xue, 2001). Following on the findings of Penner-Wilger et al. (2002), the experiment had two main theoretical purposes. The first was to provide converging evidence for the conclusion that procedure use is reflected more in τ than in μ. Penner-Wilger et al. inferred this based on expected differences in the frequencies and RT characteristics of procedures, when compared with retrieval. They had no independent way, however, to distinguish retrieval and procedure trials. Here, participants reported their strategy (retrieval or a procedure) after each trial. This allowed us to determine whether effects associated with reported procedure use appeared primarily in τ, in μ, or in both. If Penner-Wilger et al. were correct to attribute differences in τ to differences in procedure use, then effects of procedure use should appear primarily in τ rather than in μ. Demonstrating an association between reported procedures and τ would also reinforce the veridicality of the strategy reports.

The second purpose was to pursue converging evidence for the hypothesis that retrieval processes are less efficient with written word stimuli as opposed to with digit stimuli. In theory, lower retrieval efficiency is expressed in RT in two ways: (1) increased use of relatively slow procedures with the word format, and (2) slower retrieval times with the word format, especially for larger, more difficult problems (Campbell et al., 2004). We expected an association between procedure use and τ, but in contrast, we expected slower retrieval processes for word problems than for digit problems to be expressed primarily in μ rather than in τ. Retrieval trials normally constitute the majority of trials when educated adults solve simple addition problems (Campbell & Xue, 2001; LeFevre, Sadesky, & Bisanz, 1996). Consequently, a global slowing of retrieval with word stimuli ought to produce a rightward shift of the main body of the word-format distribution (i.e., μ) relative to the digit distribution. Theoretically, word-format costs in retrieval time are greater for large problems (Campbell et al., 2004); consequently, this shift in μ ought to be greater for large problems than for small problems. Thus, a format × size interaction in retrieval time should produce a format × size interaction in μ, but not necessarily in τ.

METHOD

Participants

Fifty-nine volunteers (45 women and 14 men) participated to fulfill a research option in their introductory psychology course at the University of Saskatchewan. The experiment was described on the participant sign-up sheet as investigating processes of simple calculation. Ages ranged from 17 to 26 years old (M = 19.1). The participants reported normal or corrected-to-normal vision.

Apparatus, Stimuli, and Design

Problems were displayed on a high-resolution monitor controlled by a PC. The participants wore a lapel microphone that triggered a voice-activated relay switch that was connected to the computer through the serial port. The relay switch provided the stop signal for a software clock accurate to ±1 msec.

The design of the experiment was a 2 (format: digits, words) × 2 (problem size: small, large) repeated measures design. The experiment was a replication of Experiment 2 in Campbell et al. (2004), differing only in that the participants here received twice as many blocks of trials. (A large number of observations is required to obtain reliable estimates of μ and τ.) Stimuli were the addition problems from 2 + 2 to 9 + 9. Problem operands appeared as Arabic digits or as lowercase English words, which were displayed horizontally using white characters against a dark background. The two operands in a problem were separated by the operation sign (+) and 3 spaces on each side of the operation sign, in the case of digit problems, or 1 space on each side, in the case of word problems. Therefore, digit problems occupied 8 character spaces, and the length of word problems ranged from 8–13 spaces.

The participants received eight blocks of 72 trials, with format (i.e., Arabic digits or lowercase English words) alternating across trials. There are 36 addition combinations involving the operands 2–9 when commuted pairs (e.g., 5 + 8 and 8 + 5) are counted as one problem. Within each block, the participants received all 36 problems once in word format and once in digit format. Word format was used for odd-numbered trials and digit format for even-numbered trials. The set of 36 problems included 8 "ties" (e.g., 5 + 5) and 28 "non-ties." Approximately half of the non-ties were selected randomly to be tested in the first block with the smaller operand on the left. The operand order of non-tie problems then alternated across the blocks for each operation. Problem order in each block was pseudorandom, with the constraint that word and digit versions of the same problem were separated by at least 18 trials.

To operationalize problem size, *small problems* were defined as those in which the product of the two operands was smaller than or equal to 25, and *large problems* were defined as those in which the product of the two operands was larger than 25 (Campbell, 1997; Campbell & Xue, 2001). We chose these definitions because they produced sets of small and large problems that each contained 28 non-tie problems. Tie problems were excluded from analysis on the chance that encoding-time differences between tie and non-tie problems might differ for digits and words (Campbell, 1994; Noël et al., 1997).

Procedure

Testing occurred in a quiet room with an experimenter present and lasted approximately 45 min. The addition task was preceded by a 20-trial naming task that alternated Arabic digits and English number words for naming. This allowed the participants to find a comfortable viewing distance and get accustomed to rapid responding, and it allowed the experimenter to make adjustments to the sensitivity of the voice-activated relay. For the addition task, the participants were asked to respond accurately and quickly, but were told that occasional errors are normal. The experimenter pressed a key to initiate each block. Before the first block of trials, the following instructions appeared on the monitor and were read out loud by the experimenter:

> After each problem please indicate how you solved the problem by choosing from among the following strategies: Transform, Count, Remember, Other. Say TRANSFORM if you used knowledge of a related problem. Say COUNT if you used a strategy based on counting. Say REMEMBER if the answer seemed to come to you without any intermediate steps, inferences, or calculations. Choose OTHER if you used some other strategy or are uncertain.

For reference during arithmetic trials, the participants also received a sheet of strategy descriptions, as follows:

> Transform: You solve the problem by referring to a related problem in the same or another operation. For example, you might solve 17 − 9 = ? by remembering that 17 − 10 = 7 so 17 − 9 must equal 8.
>
> Count: You solve the problem by counting a certain number of times to get the answer.
>
> Remember: You solve the problem by just remembering or knowing the answer directly from memory without any intervening steps.
>
> Other: You may solve the problem by a strategy unlisted here, or you may be uncertain how you solved the problem.

Prior to each arithmetic trial, a fixation dot appeared and flashed twice over a 1-sec interval at the center of the screen. The problem appeared (synchronized with the monitor's raster scan) on what would have been the third flash with the operation sign (+) at the fixation point. Timing began when the problem appeared, and ended when the sound-activated relay was triggered. Triggering the relay caused the problem to disappear immediately. This allowed the experimenter to mark RTs spoiled because the microphone failed to detect the onset of the response. Immediately after the response, the prompt *Strategy Choice* appeared at the center of the screen, with the words *Transform*, *Count*, *Remember*, and *Other* centered immediately below. The four words always appeared in the same order, separated by six spaces. The experimenter recorded the strategy reported by pressing one of four buttons on the computer keyboard. After the strategy had been recorded, and the experimenter had entered the stated arithmetic answer, the screen cleared and displayed the fixation dot for the next trial. No feedback about speed or accuracy was provided during the experiment.

RESULTS AND DISCUSSION

A primary goal of this experiment was to estimate μ and τ for the RT distributions associated with retrieval trials (i.e., for trials on which participants selected the remember strategy). For 12 participants, the RTSYS program (Heathcote, 1996) could not compute an estimate of μ or τ for retrieval trials in at least one format × size cell because of low rates of reported retrieval (<40 retrieval trials in a given cell). Goodness of fit for the ex-Gaussian distributional model was determined for each of the remaining participants in each condition using a chi-square statistic, with significant values indicating a poor model fit. Chi-square statistics were significant for 3 of the 47 participants ($p < .05$). Ex-Gaussian measures for these 3 participants were not excluded from the analyses because the values still provided meaningful quantitative information about the general shapes of the distributions (Penner-Wilger et al., 2002). Consequently, the results reported are based on 47 participants.

Errors

The mean error rate was 1.7%. A format × size ANOVA showed that errors were more frequent for large problems (2.4%) than for small problems (1.0%) [$F(1,46) = 26.06$, $MS_e = 3.42$, $p < .001$], and the format × size interaction replicated previous research (e.g., Campbell, 1994; Noël et al., 1997); specifically, there was a larger effect of problem size on word-format errors (+1.8%) than on digit-format errors (+0.95%) [$F(1,46) = 4.22$, $MS_e = 2.04$, $p = .046$]. Because the error rate was low and presented the usual pattern, we do not present more detailed error analyses. Our primary focus was strategy use and characteristics of the RT distributions for word and digit problems.

Strategy Reports

Most of the participants ($n = 30$) selected all four strategy categories at least once, 12 reported using three different strategies, and 5 reported using only one or two different strategies. On average, the participants reported using 3.5 different strategies during the experiment. The participants selected the remember strategy most often (71% of trials), followed by count (16%), transform (12%), and other (1%). Campbell et al. (2004, Experiment 2) used a similar experimental procedure and similarly observed the remember strategy reported most often (61% of trials), followed by count (25%), transform (13%), and other (1%). The higher rate of retrieval (i.e., remember) in the current experiment in part reflects the exclusion of participants with little reported use of retrieval in at least one format × size cell. More generally, however, participants here received each problem twice as often as in the Campbell et al. (2004) study, and therefore had more opportunities per problem to develop memory strength sufficient to use retrieval.

Table 1 presents the mean percentages of remember (i.e., retrieval), transform, count, and other strategies as a function of format and problem size. Corresponding ANOVAs of each strategy appear in Table 2. The four strategy analyses cannot be interpreted independently, of course, but it is useful to know where significant effects existed for each strategy. Retrieval was reported much less often for large problems (58% of trials) than for small problems (84% of trials), which replicates several previous studies (Campbell et al., 2004; Campbell & Xue, 2001; Hecht, 1999; LeFevre, Sadesky, & Bisanz, 1996). Large-number problems are encountered less frequently and therefore have low memory strength when compared with small problems (Ashcraft & Christy, 1995; Geary, 1996; Hamann & Ashcraft, 1986). The lower rate of retrieval for large problems mainly reflected more transformation strategies for large problems (23%) than for small problems (2%). Participants also reported counting slightly more often for large problems (18%) than for small problems (13%).

With respect to effects of format, participants reported retrieval substantially less often with problems in word format (64%) than with those in digit format (77%). The lower rate of retrieval with the word format mainly reflected more use of counting for word problems (21%) than for digit problems (11%), whereas transformation was only slightly more common for word problems (14%) than for digit problems (11%).

Table 1
Strategy Reports as a Function of Format and Problem Size

	Problem Size		
Format	Small	Large	PSE
	%Remember		
Words	78.4	50.0	28.4
Digits	88.9	65.9	23.0
W − D	−10.5	−15.9	5.4
	%Transform		
Words	2.5	24.7	−22.2
Digits	1.5	20.5	−19.0
W − D	1.0	4.2	−3.2
	%Count		
Words	18.0	23.6	−5.6
Digits	8.8	12.5	−3.7
W − D	9.2	11.1	−1.9
	%Other		
Words	1.1	1.8	−0.7
Digits	0.8	1.1	−0.3
W − D	0.3	0.7	−0.4

Note—%Count, %Remember, %Transform, and %Other, mean percentage of each strategy type reported; PSE, problem-size effect; W − D, Words minus Digits.

Table 2
Format × Size ANOVAs of Mean Percentage Reported Use of Remember, Count, and Transformation Strategies

	F			
Source	%Remember	%Count	%Transform	%Other
Format (F)	46.79^{***}	28.99^{***}	14.15^{***}	6.87^{*}
MS_e	175.11	166.24	21.68	1.87
Size (S)	76.76^{***}	3.51	48.72^{***}	10.28^{**}
MS_e	404.50	285.68	409.51	1.07
F × S	14.77^{***}	1.90	6.08^{*}	0.78
MS_e	22.50	20.54	19.45	1.58

Note—%Remember, %Count, %Transform, and %Other, mean percentage of each strategy type reported. Degrees of freedom are $F(1,46)$ for all tests. The corresponding means appear in Table 1. $^{*}p < .05$. $^{**}p < .005$. $^{***}p < .001$.

There was also a format × size interaction on reported use of retrieval; specifically, the reduction in retrieval usage with words, when compared with digits, was greater for large problems (−16%) than for small problems (−11%).[1] This interaction mainly reflected a trade-off with transformation (see Table 1). Campbell et al. (2004) observed a main effect of format on retrieval usage for addition production, but no format × size interaction. This difference between the two experiments was related to our exclusion of low-retrieval participants: When all 59 participants were included, the format × size interaction in the analysis of percentage retrieval was not significant. For individuals with relatively low retrieval strength, format evidently has a more uniform effect on retrieval usage for small and large problems. The present results more closely resemble Campbell and Fugelsang's (2001) findings for addition verification, which demonstrated a similar format × size interaction on retrieval use.[2]

In summary, use of retrieval was lower for large problems than for small problems, and this mainly reflected more use of transformation strategies for large problems. Retrieval was lower with word format than with digit format, and this mainly reflected more counting for word problems than for digit problems. The selective effects of problem size and format on transformation and counting strategies, respectively, replicate Campbell et al. (2004). To explain this dissociation, Campbell et al. (2004) proposed that over the long-term learning history for simple addition, people acquire a repertoire of transformation strategies to solve low retrieval-strength addition problems (e.g., those involving large addends). Transformation strategies typically decompose a problem into two or more subproblems involving well-known number facts (e.g., $8 + 6 = 8 + 8 = 16 - 2 = 14$). The word format, however, can impair retrieval of usually accessible facts, which would prohibit transformation strategies and leave participants to fall back on more primitive counting strategies.

We next turn to the analyses of the RT distributions for digits and words to test predictions about effects of format, problem size, and strategy use on μ and τ.

Vincentized Group RT Distributions

To graphically represent the RT distributions, group distributions were derived using RTSYS (Heathcote, 1996) for each format × size cell based on the Vincentizing technique described by Ratcliff (1979). *Vincentizing* is a means of generating an average distribution that maintains the shape of individual RT distributions. For each individual, RTs are ranked and quantiles are identified to divide the distribution into intervals that contain equal proportions of RTs. Mean RTs for corresponding quantiles are then averaged across individuals. The Vincent histogram is plotted as a series of rectangles with widths that span adjacent average quantiles and heights that produce rectangles of equal area. The area corresponds to the probability of an RT falling in that interquantile range.

Figure 1 presents the Vincent histogram for correct trials for each format × size cell. To visually represent the contribution of procedure trials to the distributions, the grayed-in portions correspond to the proportion of reported procedures associated with each area. The solid curves represent the ex-Gaussian fits to each distribution.

As can be seen in Figure 1, the digit distributions (left-hand panels) are more peaked, with less positive skew when compared with the word distributions (right-hand panels). As problem size increases (from the top to bottom plots), the digit distribution shows both a shift in the main body of the distribution (i.e., the leading edge of the normal component) and an increase in the size of the tail. The word distributions also show both a shift in the leading edge and an increase in the size of the tail. The leading edge of the distribution, however, shifts much farther for the word format.

As Figure 1 shows, procedure RTs (gray areas in Figure 1) contributed directly to these differences. As expected, procedure RTs more heavily populated the upper as opposed to the lower halves of the distributions, which is consistent with previous evidence that procedures are relatively slow compared to retrieval (see, e.g., Campbell & Xue, 2001). On average, procedure trials were 410 msec slower than were retrieval trials, which closely matched the corresponding 395-msec difference observed by Campbell et al. (2004, Experiment 2).[3] Consequently, greater procedure use for large problems and the word format would produce longer mean RTs for these conditions. Figure 1 suggests that procedure RTs also accounted substantially for the greater rightward skew for large problems, when compared with small problems, and for the greater rightward skew for word RTs relative to digit RTs. Nonetheless, for large problems in word format, with 50% reported procedure use, procedure RTs substantively

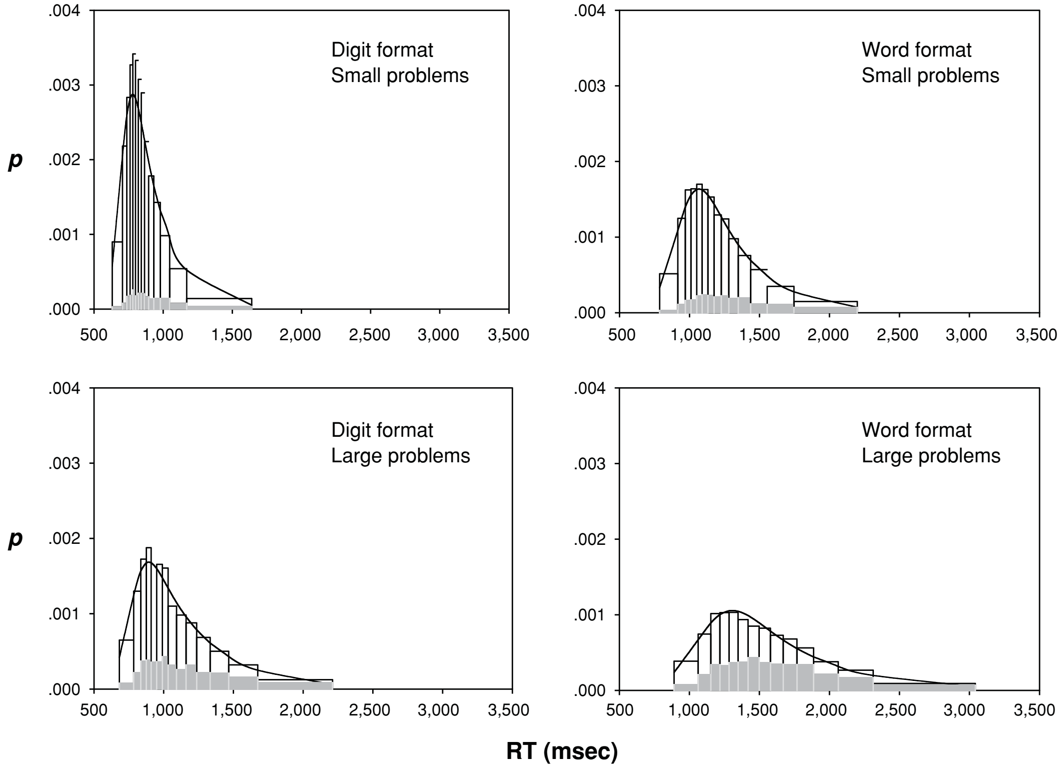

Figure 1. Group probability density histograms for distributions of RTs to digit and word format for small and large problems. Solid lines represent the fitted ex-Gaussian distribution functions. Filled portions (gray) represent proportion of procedure trials.

populated nearly the entire range of the distribution. These features of Figure 1 generally confirmed our expectations about the contribution of procedure trials to the distributions as a function of format and problem size. The following analyses precisely quantify these observations in terms of μ and τ.

RT

Figure 2 presents mean correct RT decomposed into τ and μ components (i.e., mean RT $= \mu + \tau$). The figure includes means based on all trials, regardless of strategy reported, and means for trials reportedly solved by retrieval. Mean τ corresponds to the height of the hatched, upper segment of each bar. Mean μ corresponds to the solid, lower segment of each bar. A format × size ANOVA of mean RT (all trials) confirmed the standard PSE and format effects; specifically, participants were 297 msec slower to solve large problems (1,382 msec) than to solve small problems (1,085 msec) [$F(1,46) = 91.02$, $MS_e =$ 45,437.62, $p < .001$], and 411 msec slower to solve word problems (1,439 msec) than digit problems (1,029 msec) [$F(1,46) = 238.19$, $MS_e =$ 33,306.21, $p < .001$]. Mean RT also presented a robust format × size interaction that replicated the pattern observed in previous experiments (Campbell, 1994; Campbell et al., 2004); specifically, the PSE for words (352 msec) was 110 msec larger than the PSE for digits (242 msec) [$F(1,46) = 44.32$, $MS_e =$ 3,219.00, $p < .001$].

To pursue the hypotheses presented earlier, we analyzed τ and μ with factors of format (digits or words) and problem size (small or large) separately for all trials and retrieval trials (see Figure 2). Differences between all trials and retrieval trials represent the contribution of procedure use to τ_{all} and μ_{all} (i.e., differences between the all-trials and retrieval-trials means must be attributable to the influence of trials reportedly solved by a procedural strategy).

Tau

Mean τ appears as the hatched segment of each bar in Figure 2. Overall, mean values of τ_{ret} were 50% lower than τ_{all} (means of 231 msec vs. 348 msec, averaged over the four format × size cells). Thus, the influence of procedure use on τ contributed to an overall slowing of 117 msec, when compared with retrieval.

Problem size affected both τ_{all} [$F(1,46) = 49.19$, $MS_e =$ 29,749.76, $p < .001$] and τ_{ret} [$F(1,46) = 21.33$, $MS_e =$ 8,136.31, $p < .001$], but the PSE in τ_{all} was 176 msec, compared with 61 msec in τ_{ret}. Therefore, the influence of procedure use on τ contributed 105 msec to the PSE. We show in the following section that procedure use contributed little to the PSE in μ. Thus, consistent with the conclusions of Penner-Wilger et al. (2002), the comparison of all trials and retrieval trials confirmed that procedures contributed to the PSE by influencing τ rather than μ.

Word-format costs were 127 msec for τ_{all} [$F(1,46) =$ 98.93, $MS_e =$ 7,655.29, $p < .001$], compared with 71 msec for τ_{ret} [$F(1,46) = 47.81$, $MS_e =$ 4,974.41, $p < .001$]. This implies that procedure trials contributed 56 msec to word-format costs by influencing τ. This was expected, given that procedures contributed primarily to τ, and that words led to a higher rate of procedures than did digits. Thus, the strategy reports and analysis of τ converged to support the hypothesis that the word format increased procedure use.

There was no suggestion of a format × size interaction in either τ_{all} [$F(1,46) = 0.56$, $MS_e =$ 5,921.67] or τ_{ret} [$F(1,46) = 0.30$, $MS_e =$ 5,388.93]. It follows that the format × size interaction in mean RT was associated with μ. The analysis of μ confirmed this.

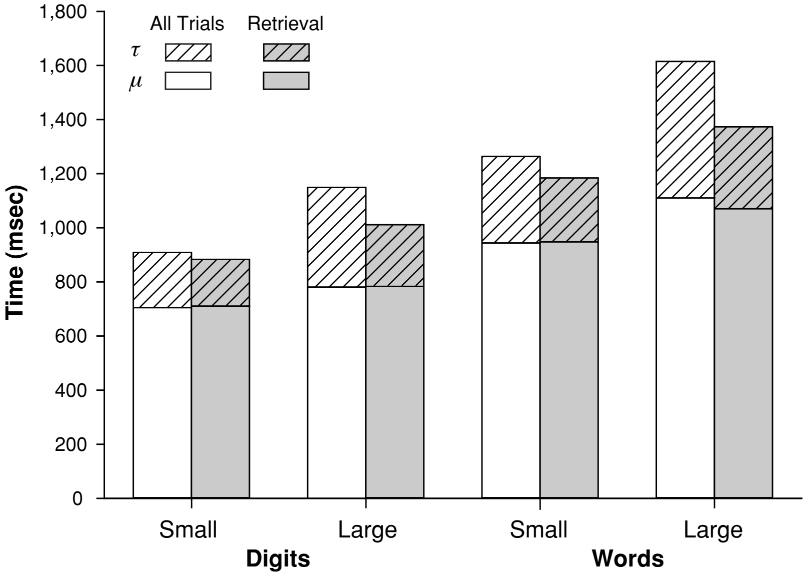

Figure 2. Mean RT for all trials and retrieval trials as a function of problem size (small, large) and format (digits, words). τ, hatched, upper segment of each bar. μ, solid, lower segment of each bar. Mean RT $= \mu + \tau$.

Mu

In Figure 2, mean μ corresponds to the solid, lower segment of each bar. The PSE was 120 msec for μ_{all} [$F(1,46) = 54.78$, $MS_e = 12{,}403.96$, $p < .001$] and 98 msec for μ_{ret} [$F(1,46) = 65.91$, $MS_e = 6{,}807.87$, $p < .001$]. Word-format costs were 284 msec for μ_{all} [$F(1,46) = 156.63$, $MS_e = 24{,}137.27$, $p < .001$] and 261 msec for μ_{ret} [$F(1,46) = 183.30$, $MS_e = 17{,}505.35$, $p < .001$]. As the figure shows, however, there was little overall difference between mean values of μ_{all} and μ_{ret}. Therefore, contrary to the findings for τ, procedure use did not generally affect μ. This confirms that procedure use expressed itself primarily in τ rather than in μ (cf. Penner-Wilger et al., 2002).

Consistent with the hypothesis that the format × size interaction reflects a disruption of retrieval, the format × size interaction was statistically robust in μ, both for all trials [$F(1,46) = 16.29$, $MS_e = 6{,}234.10$, $p = .007$] and retrieval trials [$F(1,46) = 7.93$, $MS_e = 3{,}411.65$, $p < .001$]. Specifically, for all trials, word-format costs in μ were 330 msec for large problems compared with 237 msec for small problems, corresponding to a format × size interaction effect of 93 msec. For retrieval trials, word-format costs were 285 msec for large problems, compared with 237 msec for small problems, yielding an interaction effect of 48 msec. As Figure 2 shows, the 45-msec difference between the interaction effects for all trials and retrieval trials was attributable entirely to the large, word-format condition. This implies that procedure trials contributed to the format × size interaction specifically by increasing μ in connection with large, word-format problems. Why would procedures, which were primarily associated with τ, affect μ only for large, word-format problems? This occurred because when procedures and retrieval trials contribute equally to the distribution (as in the large, word-format condition), procedures contribute RTs to the main body of the RT distribution rather than loading disproportionately in the upper tail (see Figure 1). The high proportion of procedures for large, word-format problems (50%) therefore resulted in procedure trials contributing to μ selectively in this cell of the design.

Relation Between Tau and Procedure Use

Finally, if values of τ_{all} were substantially influenced by procedure use, then we would expect a strong relation between τ_{all} and the percentage of reported procedures. In contrast, we would expect no such relation between τ_{ret} and procedure use. Indeed, the linear correlation between τ_{all} and the percentage of procedure use across the 188 format × size × participant cells was $r(186) = .42$, $p < .001$, demonstrating a strong statistical relation between τ_{all} and procedure use. As expected, there was no evidence of a statistically significant correlation between τ_{ret} and procedure use [$r(186) = .05$, $p < .05$].

DISCUSSION

We had two main theoretical goals. The first was to provide converging evidence for the conclusion that procedure use is reflected more in τ than in μ (Penner-Wilger et al., 2002). Penner-Wilger et al. found that the PSE for their China-educated group was reflected only in μ, whereas the PSE for their Canada-educated group was reflected in both μ and τ. They argued that this difference corresponded to exclusive use of retrieval by the Chinese group and a mix of solution methods by the Canadian sample. This conclusion was based on expected differences in the frequencies and RT characteristics of procedures, when compared with retrieval, but they had no independent way to distinguish retrieval and procedure trials. Here, the analyses of Figure 2 isolated the contribution of procedure trials by contrasting all-trials means and retrieval-trials means. The results confirmed that procedure trials contributed to the PSE primarily by influencing τ rather than μ. Procedures influenced τ more than μ because procedure RTs constituted a relatively small proportion of trials overall (29%), and procedure trials were 410 msec slower, on average, compared with retrieval trials; consequently, procedures were associated primarily with an upward skew (i.e., τ) rather than with a global upward shift of the distribution (i.e., μ). Thus, our results confirmed an association between reported procedures and τ, and also reinforced the veridicality of the strategy reports.

The second purpose of this experiment was to pursue evidence that retrieval is less efficient with word stimuli than with digit stimuli. Specifically, slower retrieval processes for word problems than for digit problems should be expressed primarily in μ rather than in τ: Because retrieval is the dominant strategy for adults' simple addition, a global slowing of retrieval with word stimuli ought to produce a rightward shift of the main body of the word-format distribution (i.e., μ) when compared with the digit-format distribution. Furthermore, word-format costs in retrieval time are expected to be greater for large problems (Campbell et al., 2004); consequently, the shift in μ ought to be greater for large problems than for small problems. Therefore, a format × size interaction in retrieval time should produce a format × size interaction in μ but not in τ. The analyses of Figure 2 confirmed a robust format × size interaction in μ_{ret} but no evidence of a format × size interaction in τ_{all} or τ_{ret}.

We also expected lower retrieval efficiency to be expressed in increased use of procedures for the word format. The results confirmed both a main effect of format and a format × size interaction in reported retrieval usage whereby the reduction in retrieval usage with words, as opposed to digits, was greater for large problems (−16%) than for small problems (−11%). Given that there was a format × size interaction in procedure usage (i.e., procedure use is the mirror of %Remember in Table 1), and that procedure use mainly contributed to τ, then why did we observe a format × size interaction in μ and not in τ? As discussed previously, procedure trials will load exclusively on τ only when they constitute a relatively small proportion of RTs located predominantly in the upper tail of the distribution. The format × size interaction in procedure usage occurred precisely because there was an especially high proportion of procedure trials (50%) for large, word-format problems. This high proportion resulted in procedure use for large, word-format problems contributing to μ, which localizes the format × size interaction because of procedures in μ rather than τ.

The format × size interaction in retrieval usage converges with the format × size interaction in retrieval RT to support the hypothesis that retrieval is disrupted by the word format, but especially so for the larger, more difficult problems (Campbell, 1994; Campbell & Fugelsang, 2001; Campbell et al., 2004). The hypothesis that the format × size interaction in RT might arise entirely in encoding processes (McCloskey et al., 1992; Noël et al., 1997) is contradicted by the present data (see also Campbell, 1999; Campbell & Epp, 2005). The encoding theory is mute with respect to effects of format on strategy, but our results show that the format × size interaction in retrieval usage contributed directly to the corresponding RT interaction by populating the large, word-format cell with a high proportion of the relatively slow procedural-strategy trials.

Campbell and Epp (2005) proposed that format effects on retrieval occur because problem encoding and retrieval processes for simple arithmetic are interactive rather than strictly additive stages. Format- and task-specific practice creates integrated encoding-retrieval processes that facilitate skilled numerical performance (cf. Damian, 2004). Encoding-retrieval integration entails strengthening both excitatory and inhibitory associations in order to maximize activation of relevant information and minimize activation of irrelevant information. Consequently, number-fact retrieval with problems presented in the well-practiced Arabic format is generally more skilled and efficient than when problems appear in an unfamiliar visual format, such as with written words. This effect is exaggerated for problems with relatively poor memory strength (e.g., large-number problems), because retrieval of weak associations is more disrupted by reduction in relevant information (e.g., weaker or slower activation of arithmetic associations or semantic information with word problems) or by increased interference from irrelevant processes activated by the stimulus (e.g., stronger reading-based interference with word problems; Campbell, 1994; Campbell et al., 1999).

Conclusions

Our experiment and analyses demonstrate the potential value of decomposing RT distributions to identify effects associated with μ and τ. Here, we could exploit this potential because there were clear empirical and theoretical grounds for linking arithmetic retrieval processes with μ and procedural strategies with τ, particularly when procedural strategies are relatively infrequent. The analyses of μ and τ confirmed the expected relation between RT components and reported strategy use. This reinforces our confidence that the strategy-selection method generally provides a reliable means to distinguish adults' use of retrieval from use of procedural strategies for simple arithmetic, and that analysis of strategy use can provide important insights into the cognitive architecture that mediates numerical skills.

REFERENCES

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. *Journal for Research in Mathematics Education*, **26**, 396-421.

Blankenberger, S., & Vorberg, D. (1997). The single-format assumption in arithmetic fact retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 721-738.

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

Campbell, J. I. D. (1997). Reading-based interference in cognitive arithmetic. *Canadian Journal of Experimental Psychology*, **51**, 74-81.

Campbell, J. I. D. (1999). The surface form × problem-size interaction in cognitive arithmetic: Evidence against an encoding locus. *Cognition*, **70**, 25-33.

Campbell, J. I. D., & Austin, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory & Cognition*, **30**, 988-994.

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. 457-492). Amsterdam: Elsevier.

Campbell, J. I. D., & Epp, L. J. (2005). Architectures for arithmetic. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 347-360). New York: Psychology Press.

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., & Gunter, R. (2002). Calculation, culture, and the repeated operand effect. *Cognition*, **86**, 71-96.

Campbell, J. I. D., Kanz, C. L., & Xue, Q. (1999). Number processing in Chinese–English bilinguals. *Mathematical Cognition*, **5**, 1-39.

Campbell, J. I. D., Parker, H. R., & Doetzel, N. L. (2004). Interactive effects of numerical surface form and operand parity in cognitive arithmetic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 51-64.

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.

Damian, M. F. (2004). Asymmetries in the processing of Arabic digits and number words. *Memory & Cognition*, **32**, 164-171.

Dehaene, S., & Cohen, L. (1995). Toward an anatomical and functional model of number processing. *Mathematical Cognition*, **1**, 83-120.

Geary, D. C. (1996). The problem-size effect in mental addition: Developmental and cross-national trends. *Mathematical Cognition*, **2**, 63-93.

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.

Hamann, M. S., & Ashcraft, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition & Instruction*, **3**, 173-192.

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, & Computers*, **28**, 427-445.

Hecht, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, **27**, 1097-1107.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **6**, 598-615.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, **69**, 382-386.

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., 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.

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, **27**, 364-382.

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

McCloskey, M., Macaruso, P., & Whetstone, T. (1992). The functional architecture of numerical processing mechanisms: Defending the modular model. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 493-537). Amsterdam: Elsevier.

McNeil, J. E., & Warrington, E. K. (1994). A dissociation between addition and subtraction with written calculation. *Neuropsychologia*, **32**, 717-728.

Noël, M.-P., Fias, W., & Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition*, **63**, 335-374.

Penner-Wilger, M., & Leth-Steensen, C. (2005). *Getting more from your data: Using the ex-Gaussian model to examine the shapes of response time distributions*. Manuscript submitted for publication.

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.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **114**, 510-532.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

Sciama, S. C., Semenza, C., & Butterworth, B. (1999). Repetition priming in simple addition depends on surface form and typicality. *Memory & Cognition*, **27**, 116-127.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Smith, D. G., & Mewhort, D. J. K. (1998). The distribution of latencies constrains theories of decision time: A test of the random-walk model using numeric comparison. *Australian Journal of Psychology*, **50**, 149-156.

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331-345). New York: Psychology Press.

NOTES

1. The format × size interaction in the analysis of %Remember was also significant using alternative operational definitions of problem size. With small problems defined as both operands ≤ 5 (cf. Campbell & Timm, 2000), the test of the interaction yielded $F(1,46) = 22.91$, $MS_e = 21.91$, $p < .001$. With *small* defined as sum ≤ 10 (cf. Campbell et al., 2004), $F(1,46) = 17.21$, $MS_e = 27.53$, $p < .001$.

2. We reanalyzed the Campbell et al. (2004) Experiment 2 strategy data and confirmed that the format × size interaction in retrieval usage was robust after exclusion of the 7 participants out of 50 who reported retrieval on less than 30% of trials overall.

3. We conducted analyses comparing performance on retrieval trials and procedure trials identical to those reported by Campbell et al. (2004, Experiment 2). These confirmed that the present experiment replicated every important feature of the results.

(Manuscript received May 24, 2004;
revision accepted for publication February 9, 2005.)