

# Transfer of training in alphabet arithmetic

Jamie I. D. Campbell<sup>1</sup> · Yalin Chen<sup>1</sup> · Kurtis Allen<sup>1</sup> · Leah Beech<sup>1</sup>

Published online: 28 June 2016  
© Psychonomic Society, Inc. 2016

**Abstract** In recent years, several researchers have proposed that skilled adults may solve single-digit addition problems (e.g.,  $3 + 1 = 4$ ,  $4 + 3 = 7$ ) using a fast counting procedure. Practicing a procedure, however, often leads to transfer of learning to unpracticed items; consequently, the fast counting theory was potentially challenged by subsequent studies that found no generalization of practice for simple addition. In two experiments reported here ( $Ns = 48$ ), we examined generalization in an alphabet arithmetic task (e.g.,  $B + 5 = C D E F G$ ) to determine that counting-based procedures do produce generalization. Both experiments showed robust generalization (i.e., faster response times relative to control problems) when a test problem's letter augend and answer letter sequence overlapped with practiced problems (e.g., practice  $B + 5 = C D E F G$ , test  $B + 3 = C D E$ ). In Experiment 2, test items with an unpracticed letter but whose answer was in a practiced letter sequence (e.g., practice  $C + 3 = D E F$ , test  $D + 2 = E F$ ) also displayed generalization. Reanalysis of previously published addition generalization experiments (combined  $n = 172$ ) found no evidence of facilitation when problems were preceded by problems with a matching augend and counting sequence. The clear presence of generalization in counting-based alphabet arithmetic, and the absence of generalization of practice effects in genuine addition, represent a challenge to fast counting theories of skilled adults' simple addition.

**Keywords** Addition · Procedures · Counting · Transfer of training

✉ Jamie I. D. Campbell  
jamie.campbell@usask.ca

<sup>1</sup> Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK, Canada S7N 5A5

There is an ongoing debate in the cognitive science literature about the processes that subserve skilled adults' simple mental addition (e.g.,  $3 + 2 = 5$ ,  $4 + 3 = 7$ ). For years, research seemed to support the theory that adults usually solve single-digit addition problems by direct retrieval from associative memory (see Ashcraft & Guillaume, 2009; Zbrodoff & Logan, 2005, for reviews). Recently, however, several researchers have argued that the counting process commonly used for addition in childhood (Groen & Parkman, 1972) evolves into an automatic or “compacted” counting procedure in highly skilled adults. The present experiments were designed to investigate if transfer of training (i.e., generalization of learning from practiced to unpracticed problems) occurs with counting-based procedures. If it does, this would reinforce the case that generalization of practice provides a marker for the use of counting procedures in mental addition.

## Evidence of fast counting procedures for simple addition by skilled adults

Fayol and Thevenot (2012) reignited the debate (Ashcraft & Fierman, 1982; Groen & Parkman, 1972) about the basic processes for skilled adults' simple addition. They used an operator-priming paradigm (see also Roussel, Fayol & Barrouillet, 2002; Sohn & Carlson, 1998) and tested engineering students in blocks of mixed simple addition and multiplication problems. When the operation sign (+ or  $\times$ ) appeared 150 ms before the problem operands, response time (RT) for addition problems was faster relative to simultaneous presentation; but there was no operator preview effect for multiplication. Fayol and Thevenot proposed that addition was solved via a fast procedure that could be primed by a preview of the plus sign, whereas multiplication involved direct

memory retrieval of individual facts and therefore was not subject to operator priming (but see Chen & Campbell, 2015).

Subsequently, Barrouillet and Thevenot (2013; see also Thevenot, Barrouillet, Castel, & Uittenhove 2016; Uittenhove, Thevenot, & Barrouillet, 2016) analyzed adults' response times (RTs) for very small additions involving the numbers 1 through 4 and found that RT increased linearly with the numerical size of the addends. They proposed that small addition problems could activate a very fast counting procedure (20 ms per step) that gives rise to a shallow linear problem size effect. Campbell, Chen, and Maslany (2013) provided another type of evidence that small addition problems may be solved by fast procedures, at least in highly skilled individuals. They examined Canadian and Chinese adults' performance in an arithmetic interference paradigm. Practicing small multiplication problems (e.g.,  $2 \times 3$ ) slowed RT to answer addition counterparts ( $2 + 3$ ) for the Canadian group but not for the Chinese group. As this retrieval-induced interference effect had previously been shown to be induced by number-fact retrieval practice but not by practice of arithmetic procedures (Campbell & Theriault, 2013), Campbell et al. suggested that the arithmetically superior Chinese participants might solve small addition problems by fast procedures whereas the Canadians used number-fact retrieval. Despite these seemingly converging sources of evidence for fast addition procedures, there appears to be little evidence outside of the addition domain to support the idea that mediational strategies for associative learning become proceduralized with practice (see, e.g., Kole & Healy, 2013; Rickard & Bajic, 2003). Furthermore, as we discuss next, repeated attempts have failed to demonstrate a basic behavioral prediction of counting procedures for simple addition.

### Generalization of addition practice

To pursue the issue of fast procedures in adults' simple addition, Campbell and Beech (2014) examined generalization of practice. Practicing a procedural process results in its speeding up (Singley & Anderson, 1989); consequently, speed up with practice should generalize to different, unpracticed problems that use that procedure. Campbell and Beech reasoned that if simple addition was based on procedures, then practicing a subset of problems (e.g.,  $4 + 3$ ) should facilitate subsequent performance of similar, unpracticed problems (e.g.,  $3 + 2$ ). The results showed that there was no generalization of practice for nonzero simple addition problems, but the procedure-based  $0 + N = N$  problems presented clear evidence of generalization (i.e., practicing a subset of  $0 + N$  problems facilitated a different subset of  $0 + N$  problems). Generalization for  $0 + N$  problems, but no generalization for nonzero simple addition problems has been repeatedly replicated (Campbell, Dufour,

& Chen, 2014; Campbell & Theriault, 2013; Chen & Campbell, 2014, 2016).

If generalization is a reliable marker of procedure use, then the null generalization results of Campbell and Beech (2014) and similar studies (e.g., Chen & Campbell, 2014, 2016) cast doubt on the general applicability of the theory of fast procedures for simple addition. There is no direct evidence however, that counting-based procedures do produce robust generalization. In simple arithmetic, generalization has only been demonstrated for problems governed by the identity rules  $0 + N = N$ ,  $1 \times N = N$ , and  $0 \times N = 0$ . The procedures underlying application of the identity rules might be different in kind from the fast counting procedure proposed by Barrouillet and Thevenot (2013). The purpose of these experiments was to find out if practice of counting-based procedures produces generalization.

To this end we used a version of the alphabet-arithmetic paradigm that Logan (1988) developed to study the automaticity of cognitive skills (see Zbrodoff & Logan, 2005, for a review of research with this paradigm). Logan and Klapp (1991, Experiment 1) assessed whether extended practice is necessary to produce automaticity. Participants were asked to verify equations with either the first 10 letters (A–J) or the second 10 letters of the alphabet (K–T) combined with the digit addends 2–5 (e.g.,  $A + 2 = C$ ,  $D + 4 = H$ ). Initially, alphabet arithmetic involves counting forward through the alphabet from the augend letter with step-by-step enumeration of successive letters. The linear RT slope as a function of addend size in the first of 12 sessions (486 ms per addend increment) provided evidence of counting to verify problems. Repeated practice, however, reduced the slope to 45 ms, which is less than the alphabet recitation time per letter recorded by the experimenters (115 ms). Note, however, that the RT slope from +2 to +4 (i.e., excluding +5) was about 120 ms per increment (Fig. 1, p. 182). Most important in the present context, little transfer was observed when participants were switched to the unstudied half of the alphabet. In Logan and Klapp's Experiment 3, participants were trained and tested on a variety of alphabet-arithmetic facts. In the transfer phase, they verified the old (i.e., trained) facts, new facts composed of the old letters (new digits, old letters), and new facts composed of new letters (new digits, new letters) to assess transfer. The addend-size slopes for both new-new and new-old problems were steep and did not differ from each other suggesting that similarity (i.e., overlapping features in trained and transfer problems) did not promote facilitative transfer. Logan and Klapp concluded that participants only learned the particular items that they practiced; consequently, transfer to new, unpracticed items was weak.

## The present experiments

For the purposes of Logan and Klapp (1991), the evidence of poor transfer of alphabet-arithmetic practice to new problems supported the theory that performance became automated by a transition during practice from a counting-based algorithm to storage and retrieval of individual facts in memory. These findings and conclusions raise a challenge for the automatic-counting theories of skilled simple addition (e.g., Fayol & Thevenot, 2012), which assume that continued practice of counting for addition during learning results in the counting algorithm becoming automatic. The findings referred to here were the lack of transfer observed, not the flat RT slope. Other lines of experimental evidence also indicate a transition from algorithmic procedures to fact retrieval as skill develops (Bajic & Rickard, 2009; Barrouillet & Fayol, 1998; Rickard, Lau, & Pashler, 2008; see also Kole & Healy, 2013). Nonetheless, the results of Logan and Klapp also raise doubts that practicing a counting algorithm necessarily leads to measurable generalization of practice to new items.

In the following studies, we examined transfer of alphabet arithmetic practice to new items after only six repetitions of each practice item. This should permit transfer to be measured while counting remained the predominant solution method during the practice phase. This is important because we wanted specifically to test transfer of counting practice to new items also solved by counting. The stimuli used are in Table 1. Two types of transfer items were created: augend-sequence and sequence-only items. Augend-sequence transfer items are composed of a practiced letter augend with an answer in the practiced letter sequence. For example,  $B + 3$  is an augend-sequence transfer item when  $B + 5$  is practiced. To solve  $B + 5$  by counting, five incremental steps through the alphabet are required: C D E F G. Thus,  $B + 3$  is an augend-sequence transfer item because it shares the augend letter B with the practiced problem and its answer (E) is in the letter counting sequence generated by solving  $B + 5$ . In contrast,  $D + 3$  is a sequence-only transfer item in relation to  $B + 5$  because its augend letter D is not practiced, but its answer (G) is in the practiced sequence. Use of the two sequence types (i.e., augend-sequence and sequence-only) allowed us to determine if generalization of practice required transfer problems to share both augend and answer-sequence features with practiced problems or if practicing an overlapping answer sequence alone would yield generalization of counting practice to test problems.

There were four practice sets of five alphabet arithmetic problems based on the successive letter pairs BC, HI, NO, and TU. Each was related to one of four test sets of six problems based, respectively, on the letter pairs BD, HJ, NP, and TV (see Table 1). In the practice problems, the augend letters B, H, N, and T were combined with the addends +4 and +5, and the letters C, I, O, and U combined with the

**Table 1** The practice and test phase alphabet–arithmetic sets used in Experiments 1 and 2

Practice Sets		Test Sets
CB		BD
$C + 1 = D$		$B + 1 = C$
$C + 2 = E$	Augend-sequence	$B + 2 = D$
$C + 3 = F$		$B + 3 = E$
$B + 4 = F$		
$B + 5 = G$	Sequence-only	$D + 1 = E$
		$D + 2 = F$
		$D + 3 = G$
IH		HJ
$I + 1 = J$		$H + 1 = I$
$I + 2 = K$	Augend-sequence	$H + 2 = J$
$I + 3 = L$		$H + 3 = K$
$H + 4 = L$		
$H + 5 = M$	Sequence-only	$J + 1 = K$
		$J + 2 = L$
		$J + 3 = M$
ON		NP
$O + 1 = P$		$N + 1 = O$
$O + 2 = Q$	Augend-sequence	$N + 2 = P$
$O + 3 = R$		$N + 3 = Q$
$N + 4 = R$		
$N + 5 = S$	Sequence-only	$P + 1 = Q$
		$P + 2 = R$
		$P + 3 = S$
UT		TV
$U + 1 = V$		$T + 1 = U$
$U + 2 = W$	Augend-sequence	$T + 2 = V$
$U + 3 = X$		$T + 3 = W$
$T + 4 = X$		
$T + 5 = Y$	Sequence-only	$V + 1 = W$
		$V + 2 = X$
		$V + 3 = Y$

*Note.* Each practice set had a test set yoked uniquely to it by a shared augend letter and answer sequence (augend-sequence) or answer sequence only (sequence-only). Participants practiced two of the practice sets for six blocks and then were tested on all of the test sets. Test problems yoked to the practiced sets were transfer problems; the others were control problems

addends +1, +2, and +3. In the test problem sets, all the letters were paired with the addends +1, +2, and +3. As a result, for each practice set (e.g.,  $C + 1 = D$ ,  $C + 2 = E$ ,  $C + 3 = F$ ,  $B + 4 = F$ ,  $B + 5 = G$ ) there were three test problems that shared both a common augend and answer sequence with the practice set ( $B + 1 = C$ ,  $B + 2 = D$ ,  $B + 3 = E$ ) and three that overlapped with the

practice set only in sharing a common answer sequence ( $D + 1 = E$ ,  $D + 2 = F$ ,  $D + 3 = G$ ). Participants were trained on two of the practice sets and tested on all four test sets. The test sets related to the practiced sets by virtue of overlapping features constituted transfer items, and other two test sets were control items.

We also investigated possible effects of English as one's first and primary reading language versus English as a secondary language. The purpose was to explore if alphabet familiarity affected transfer potential. We reasoned that RT gains during practice might be greater when alphabet fluency was initially relatively low, which could translate into greater transferred gains at test. In summary, we expected generalization of practice at test for the transfer problems relative to control problems, assuming that transfer of counting skill occurs between problems that share common procedural components, but a relatively larger transfer effect for participants with lower alphabet familiarity.

## Experiment 1

### Method

#### *Participants*

Forty-eight participants recruited at the University of Saskatchewan received either course credit or \$5 CAD. Twenty-four participants reported English as their first language (22 women, 24 right-handed, ages 17–35 years,  $M = 24.2$ ,  $SD = 5.4$ ). The remaining 24 participants identified their first language as not English (13 women, 24 right-handed, ages 19–57,  $M = 29.5$ ,  $SD = 10.2$ ), including 11 Chinese, two Vietnamese, two Hindi, and one each of Bangla, German, Hungarian, Kannada, Korean, Russian, Sinhala, Spanish, and Urdu. Participants answered the alphabet arithmetic problems in English.

#### *Stimuli*

There were four practice sets of five alphabet arithmetic problems based on the successive letter pairs BC, HI, NO, and TU. Each was related to one of four test sets of six problems based, respectively, on the letter pairs BD, HJ, NP, and TV (see Table 1). In the practice problems, the augend letters B, H, N, and T were combined with the addends +4 and +5, and the letters C, I, O, and U combined with the addends +1, +2, and +3. In the test problem sets, all the letters were paired with the addends +1, +2, and +3. As a result, for each practice set there were three related test problems that shared both a common augend and answer sequence with the practice set (augend-sequence transfer items) and three that overlapped with the

practice set only in sharing a common answer sequence (sequence-only transfer items).

Participants were trained on two of the four practice sets, one set from the first half of the alphabet (BD or HJ) and one from the second half (NP or TV). Assignment of the two practice sets was counterbalanced across groups of four participants within each language group. All four of the test problem sets were presented in the test phase. Two of these test sets served as transfer problems that shared a common augend and/or answer sequence with the practice problems, whereas the other two problems sets served as control problems with unpracticed augends and unpracticed letter answer sequences. Thus, counterbalancing of the practice sets across participants simultaneously effected counterbalancing of transfer and control problem sets. Augend-sequence and sequence-only problems were defined relative to their yoked practice set of problems; consequently, they necessarily involved different problems and could not be counterbalanced. Therefore, any overall difference in performance between augend-sequence and sequence-only problems might be attributable to intrinsic differences in item difficulty.

#### *Apparatus, design, and procedure*

The stimuli were presented on two CRT monitors controlled by E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). One monitor was viewed by the participant, and the other by the experimenter. The experimenter's monitor also showed trial and block information and a running tally of the number of errors made by the participant. Participants sat approximately 50 cm from a monitor and spoke into a microphone that they held. The verbal response triggered the stop signal to a software clock accurate to  $\pm 1$  ms.

Prior to the alphabet arithmetic task, participants were instructed to recite the alphabet aloud three times, quickly and accurately, and were timed to the nearest second. This provided a simple measure of fluency with the English alphabet sequence for the English-first and not-English-first groups. There were six alphabet arithmetic training blocks of 10 problems followed by two test blocks of 24 problems that took approximately 15 minutes to complete. For each participant within each block, problem order was independently randomized. The problems were displayed in black, Courier New 14-point font on a white background. The displayed problem occupied five character spaces with the augend and addend separated by the plus sign with adjacent spaces (e.g., B + 3). Each trial started with a 1-s central fixation dot, then the problem appeared with the plus sign at fixation. Response timing commenced with the appearance of the problem and was stopped by the participant's verbal response. Accuracy rather than speed was emphasized in the training phase, but in the test phase, participants were instructed to respond both quickly and accurately. After the verbal response was provided, the



experimenter entered the participant's answer and marked spoiled RTs when the stop signal was not activated by response onset. The fixation dot then appeared to initiate another trial. Participants were permitted a short break between the training and test phase. No feedback was given on responses for either phase.

## Results

### Alphabet recitation

A *t* test comparing the English-first and not-English-first groups' mean times for the third (i.e., final) alphabet recitation trial indicated faster mean recitation by the English-first (5.1 s) compared to the not-English-first group (8.4 s),  $t(45) = 3.52$ ,  $p = .002$ ,  $SE = .93$ .<sup>1</sup>

### Practice phase

Median RT for correct responses received a Group (2) × Block (6) ANOVA. The corresponding means appear in Table 2. Greenhouse-Geisser corrected statistics are reported when Mauchly's test indicated violation of the sphericity assumption. Despite the difference in alphabet recitation times, there were no significant differences between the groups' alphabet arithmetic RT during the practice phase (all  $ps > .13$  for group-related tests including the linear through fifth-order contrasts). Mean practice RT was 2,288 ms for the English-first and 2,311 ms for the not-English-first. Mean RT sped up by 30.3% across the six practice blocks (2,887 ms, 2,377 ms, 2,237 ms, 2,236 ms, 2,054 ms, and 2,012 ms),  $F(3.6, 166.8) = 28.75$ ,  $p < .0001$ ,  $MSE = 6,572,718$ ,  $\eta_p^2 = .39$ . With respect to errors during practice (see Table 2), English-first participants made fewer errors than not-English-first in practice Block 1 (8.3% vs. 18.3%),  $F(1, 46) = 4.43$ ,  $p = .04$ ,  $MSE = 271.01$ ,  $\eta_p^2 = .09$ , but the groups produced equivalent error rates over Blocks 2 through 6 (both 6.9%), and there were no effects of practice block on errors over the last five blocks (all  $ps > .20$ ).

We also examined effects of addend size (+1 to +5) during practice in Group (2) × Addend (5) ANOVAs of median RT and percentage of errors. Mean median RT increased as addend size increased from +1 to +5 (1,195 ms, 1,745 ms, 2,438 ms, 2,805 ms, and 3,073 ms), presenting both linear,  $F(1, 46) = 379.18$ ,  $p < .0001$ ,  $MSE = 293,439$ ,  $\eta_p^2 = .89$ , and quadratic components,  $F(1, 46) = 31.78$ ,  $p < .0001$ ,  $MSE = 87,243$ ,  $\eta_p^2 = .40$ , the latter owing primarily to an inflection point at +3. There were no effects of group (all  $ps > .30$ ). To determine if the linear and quadratic effects were present throughout practice, or emerged as practice progressed, we

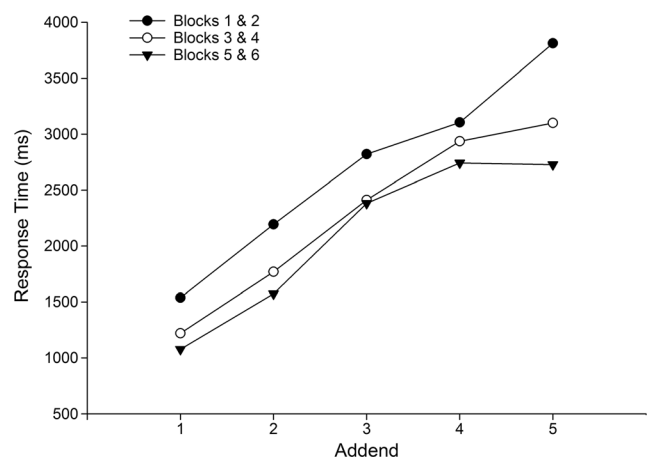
<sup>1</sup> One not-English-first participant was excluded from this analysis because no time was recorded for the third alphabet recitation trial.

**Table 2** Mean response time in milliseconds and mean percentage of errors (standard errors in parentheses) in the practice phase by group and block in Experiment 1

Block	Response Time		Error Percentage	
	English 1st	Not English 1st	English 1st	Not English 1st
1	2947 (142)	2817 (116)	8.3 (2.1)	18.3 (4.2)
2	2325 (119)	2430 (141)	7.9 (1.9)	9.2 (2.5)
3	2250 (141)	2223 (122)	6.3 (2.2)	5.8 (1.8)
4	2267 (174)	2205 (118)	5.8 (1.9)	5.8 (2.2)
5	2010 (141)	2098 (103)	7.9 (2.3)	6.7 (2.0)
6	1929 (106)	2094 (136)	6.7 (2.5)	7.1 (1.6)

calculated the median RT for each addend combining Blocks 1 and 2, Blocks 3 and 4, and Blocks 5 and 6 to represent early, middle, and later stages of practice, respectively. Each of the three practice stages received an ANOVA with addend size as a repeated-measures factor.

As Fig. 1 shows, early in practice (i.e., Blocks 1 and 2), RT increased linearly with addend size,  $F(1, 47) = 164.11$ ,  $p < .0001$ ,  $MSE = 872,528$ ,  $\eta_p^2 = .78$ , and there were no significant deviations from linearity (i.e., none of the higher order contrasts were significant (all  $ps > .07$ ;  $p = .55$  for the quadratic contrast). The linear pattern for early practice is consistent with exclusive use of a counting-based strategy with each increment through the alphabet requiring about 550 ms on average. In Blocks 3 and 4, there remained a strong linear effect of addend size,  $F(1, 47) = 288.28$ ,  $p < .0001$ ,  $MSE = 404,627$ ,  $\eta_p^2 = .86$ , but a robust quadratic component appeared,  $F(1, 47) = 15.90$ ,  $p = .0002$ ,  $MSE = 170,674$ ,  $\eta_p^2 = .25$ , owing primarily to +5 deviating below a linear position. Over the final two blocks of practice, beyond the linear effect,  $F(1, 47) = 296.49$ ,  $p < .0001$ ,  $MSE = 323,815$ ,  $\eta_p^2 = .86$ , there appeared both quadratic,  $F(1, 47) = 27.27$ ,  $p < .0001$ ,



**Fig. 1** Mean response time in the practice phase of Experiment 1 by practice blocks and addend

$MSE = 271,898$ ,  $\eta_p^2 = .37$ , and cubic,  $F(1, 47) = 9.58$ ,  $p = .0003$ ,  $MSE = 237,158$ ,  $\eta_p^2 = .17$ , trends owing primarily to +4 and +5 deviating below the linear positions expected based on mean RTs for +1 to +3. These results are consistent with +4 and +5 problems beginning a transition from counting-based solutions to memory-fact retrieval.

The Group  $\times$  Addend ANOVA of percentage of errors indicated only a linear effect,  $F(1, 46) = 17.03$ ,  $p < .0002$ ,  $MSE = 145.83$ ,  $\eta_p^2 = .27$ , of addend, with errors generally increasing as addend size increased from +1 to +5 (2.4%, 6.3%, 9.7%, 9.2%, 12.3%).

*Test phase*

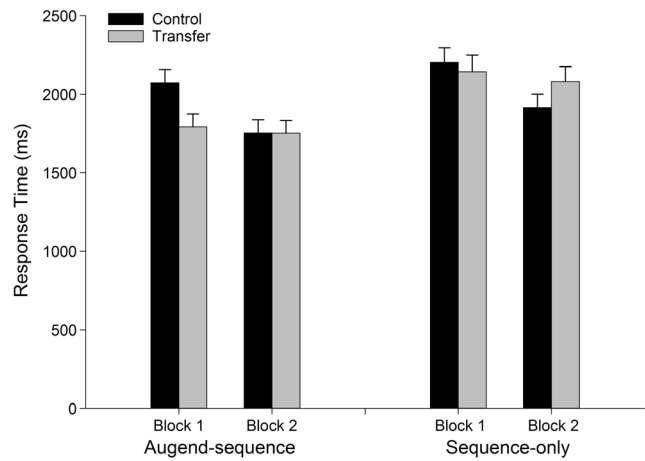
Mean median RT for correct responses confirmed a counting-based strategy for test phase problems with RT increasing from 1,399 to 1,877 to 2,484 ms for +1, +2, and +3, respectively. The corresponding error rates were 2.7%, 5.9%, and 7.0% for +1 to +3, respectively.

To evaluate transfer, median RT for correct answers received a Group  $\times$  Sequence Type (augend-sequence vs. sequence-only)  $\times$  Transfer Type (transfer vs. control)  $\times$  Block (1 vs. 2) mixed-factor ANOVA. The corresponding means appear in Table 3. The group factor did not participate in any significant effects, although the test of the main effect suggested a small overall RT advantage for the English-first group (1,837 ms) compared to the not-English-first group (2,092 ms),  $F(1, 46) = 3.43$ ,  $p = .07$ ,  $MSE = 1,826,579$ ,  $\eta_p^2 = .07$ . Figure 2 presents the mean RT as a function of sequence type, transfer type, and block averaged over the two groups.

Participants were faster overall in test Block 2 (1,876 ms) compared to Block 1 (2,053 ms),  $F(1, 46) = 24.55$ ,  $p < .0001$ ,  $MSE = 122,842$ ,  $\eta_p^2 = .35$ , but this was qualified by the Block  $\times$  Transfer Type interaction,  $F(1, 46) = 13.20$ ,  $p = .001$ ,  $MSE = 116,574$ ,  $\eta_p^2 = .22$ . As Fig. 2 shows, this occurred because transfer problems were faster than control problems in Block

**Table 3** Mean response time in milliseconds and standard error (in parentheses) in the test phase by group, block, problem type (augend-sequence, sequence-only) and training type (control, transfer) in Experiment 1

	Augend-sequence		Sequence-only	
	Control	Transfer	Control	Transfer
English 1st				
Block 1	1880 (96)	1711 (118)	2010 (110)	2058 (138)
Block 2	1502 (59)	1698 (119)	1831 (85)	2005 (156)
Not English 1st				
Block 1	2268 (139)	1873 (117)	2396 (148)	2230 (158)
Block 2	2005 (159)	1808 (110)	2001 (149)	2158 (105)



**Fig. 2** Mean response time in the test phase of Experiment 1 as a function of sequence type, transfer type, and block. Error bars are  $\pm 1$  standard error

1 (170 ms),  $t(47) = 2.17$ ,  $p = .03$ ,  $SE = 678.43$ , but not in Block 2 (-83 ms),  $t(47) = 1.16$ ,  $p = .25$ ,  $SE = 71.34$ . Thus, the experiment demonstrated transfer of learning in alphabet arithmetic in Block 1. This effect disappeared in Block 2 because control problems sped up substantially relative to Block 1 when they were repeated in Block 2, whereas the transfer problems did not show speed-up across test blocks (see Fig. 2).

Apart from these effects, augend-sequence problems were on average 243 ms faster overall than sequence-only problems,  $F(1, 46) = 53.22$ ,  $p < .0001$ ,  $MSE = 106,585$ ,  $\eta_p^2 = .54$ , but there was a Sequence Type  $\times$  Transfer Type interaction,  $F(1, 46) = 4.19$ ,  $p = .046$ ,  $MSE = 216,849$ ,  $\eta_p^2 = .08$ . This interaction occurred because the RT advantage for augend-sequence problems compared to sequence-only problems was larger in the transfer condition, 340 ms,  $t(47) = 6.55$ ,  $p < .0001$ ,  $SE = 51.9$ , than in the control condition, 146 ms,  $t(47) = 2.29$ ,  $p = .03$ ,  $SE = 63.61$ . The significant RT advantage for augend-sequence problems relative to sequence-only problems in the control condition suggests the former were intrinsically easier. For example, the longer average RT for sequence-only problems might reflect them involving letters later in the alphabet relative to augend-sequence problems.

Nonetheless, the overall advantage for augend-sequence problems owed substantially to facilitative transfer observed for these items. In fact, the only statistically clear transfer effect occurred for augend-sequence problems in Block 1, 282 ms,  $t(47) = 2.74$ ,  $p = .009$ ,  $SE = 102.8$ , whereas sequence-only problems did not present evidence of positive transfer in Block 1, 59 ms,  $t(47) = 0.60$ ,  $p = .55$ ,  $SE = 98.5$ .

Percentage of errors during the test phase (5.2% overall) also received a Group  $\times$  Sequence Type  $\times$  Transfer Type  $\times$  Block mixed-factor ANOVA (see Table 4 for the corresponding means and standard errors). The only significant effect was the Sequence Type  $\times$  Transfer Type  $\times$  Block interaction,  $F(1, 46) = 7.55$ ,  $p = .009$ ,  $MSE = 129.58$ ,  $\eta_p^2 = .14$ . For augend-

**Table 4** Mean error rate and standard errors (in parentheses) in the test phase for group, block, problem type (augend-sequence, sequence-only) and training type (control, transfer) in Experiment 1

	Augend-sequence		Sequence-only	
	Control	Transfer	Control	Transfer
English 1st				
Block 1	4.2 (1.8)	4.9 (2.1)	5.6 (1.9)	6.3 (2.2)
Block 2	6.9 (2.4)	2.1 (1.1)	4.2 (1.5)	7.6 (2.0)
Not English 1st				
Block 1	5.6 (1.9)	6.9 (2.6)	9.0 (2.8)	4.9 (1.9)
Block 2	4.9 (2.1)	2.8 (1.9)	2.8 (2.1)	4.9 (1.6)

sequence problems there were fewer errors for transfer than control items in Block 2 (-3.5%) but not in Block 1 (+1.0%), whereas sequence-only problems presented the reverse pattern (+2.8% in Block 2 vs. -1.7% in Block 1). We did not attempt to interpret these small differences.

## Discussion

RTs for the augend-sequence problems indicated robust transfer of practice in Block 1. This confirms that practice of counting-based processes can yield generalization effects that facilitate related, unpracticed problems. There was no significant transfer for sequence-only problems, however, which suggests that the matching augend feature of augend-sequence problems was necessary for robust transfer in Experiment 1. The transfer effect was observed in Block 1 but not in Block 2. This occurred because control problems sped up substantially relative to Block 1 when they were repeated in Block 2 (304 ms), but transfer problems did not present substantial speed-up across test blocks (50 ms; see Fig. 2). The 304 ms of speed-up observed for control problems was similar to speed-up during practice for +1 to +3 problems from Block 1 to 2 (377 ms). For the augend-sequence problems, these asymmetrical repetition effects would owe, at least in part, to the substantial transfer of learning from the practice phase, which would limit further gains observed in Block 2 from a single repetition of these problems in Block 1 of the test phase.

Nonetheless, the sequence-only transfer problems did not present significant generalization of learning in Block 1, but still showed little benefit from repetition in the test phase compared to the sequence-only control items. This suggests that although no significant generalization occurred for sequence-only transfer items, practice of their answer sequences limited potential gains from repeating these problems in the test phase. One possibility is that the RT gains from repetition for control problems reflect use of episodic short-term memory in Block 2 to directly recall the problem solving

episode for specific problems from Block 1 and thereby sometimes bypass slower counting-based processes. For the transfer problems, exploitation of episodic memory might be more difficult because their answers were repeatedly associated with the practiced problems, which would introduce associative interference with memory and limit use of retrieval for transfer problems. In Experiment 2, we introduced a backward-counting task between blocks to reduce potential contributions of episodic short-term memory.

Early in the practice phase (Blocks 1 and 2) RT was strictly linear as a function of addend size from +1 to +5. The linear pattern suggests exclusive use of a counting-based strategy for all practice problems with each step through the alphabet requiring about 550 ms. In Blocks 3 and 4, however, a quadratic component appeared owing primarily to +5 deviating below a linear position. Over the final two blocks of practice, there were both quadratic and cubic components owing primarily to +4 and +5 deviating below the linear positions expected based on mean RTs for +1 to +3.

The relatively greater gains for +4 and +5 across practice blocks could reflect a fan effect because the augend letters for +4 and +5 problems were each associated with two problems, whereas +1, +2, and +3 problems were each associated with three problems. This constitutes a manipulation of associative fan (e.g., Anderson & Reder, 1999; Pirolli & Anderson, 1985), with +4 and +5 having smaller associative fan than +1 to +3, and therefore less potential associative interference from related items. This would promote memorization of +4 and +5 items relative to the small-addend problems and lead to relative RT gains because an increasing percentage of trials would be based on relatively fast fact retrieval in place of multistep counting. Such a fan effect has been observed previously with alphabet arithmetic stimuli (Zbrodoff, 1995). Nonetheless, Logan and Klapp (1991, pp. 182–183) attributed a similar discontinuity for +5 alphabet arithmetic to a shift from counting to memory-based performance, although fan was not manipulated in their experiments. Perhaps because 5 is the maximum addend, +5 problems are distinctive and promote early development of mnemonic strategies for these items.

Finally, although the English-first group recited the English alphabet faster than the not-English-first group, there was little evidence that this difference mattered for the alphabet arithmetic task, and there was no evidence that practice or transfer effects differed between the groups. Both groups apparently had sufficient familiarity and access to the alphabet sequence to use it similarly and perform alphabet arithmetic at the same level. As the first-language manipulation had no apparent effects on alphabet arithmetic performance, we did not pursue it in Experiment 2.

## Experiment 2

The only significant procedural difference between Experiment 1 and Experiment 2 was the introduction of a backward-counting task between blocks. At the end of each block, participants immediately saw a three-digit number and were instructed to count backwards by threes until the cue for the next trial appeared 10 seconds later. This modification was motivated by two considerations. First, during the practice phase in Experiment 1, there were no planned breaks between blocks and participants usually proceeded immediately to the next block. As a result of this, participants sometimes would have been able to use episodic or short-term memory to directly retrieve a recent previous episode with a given problem and bypass counting. This would work against finding transfer in the test phase because it is specifically counting-related transfer that we attempted to measure. The 10-second interpolated counting task would flush working memory contents and prevent possible memory rehearsal of previous items between blocks.

Second, during the test phase in Experiment 1, control problems presented substantial speed up across blocks relative to transfer problems. For control items, it might have been relatively easier to use episodic short-term memory from Block 1 to solve some problems in Block 2, whereas episodic memory for transfer items would encounter interference from the practice of related problems during the training phase. The countdown task potentially could reduce the contribution of episodic retrieval strategies in Block 2 of the test phase.

## Method

Forty-eight participants who had not participated in Experiment 1 were recruited at the University of Saskatchewan and received course credit or \$7.50 CAD. The increase in monetary compensation relative to \$5.00 in Experiment 1 reflected new participant compensation standards. There were 13 men and 35 women with a mean age of 20.5 years ( $SD = 3.24$ ). Forty-four were right-handed and four were left-handed. Thirty-six participants reported English as their first language, and 12 reported English as not first, including eight Chinese, two Arabic, and one each of French and Hindi.

The stimuli, design, apparatus and procedure were the same as Experiment 1, except for the following things. First, to check the sensitivity of the microphone, participants performed an eight-trial word-naming task before experimental trials. Second, to reduce potential use of short-term memory or episodic memory across successive blocks, a 10-second count-down task was performed by the participant after each practice block and between the two test phase blocks. Specifically, participants were required to count backwards by threes from the number 100 plus the block number (101, 102, 103, etc.). The starting number for the countdown was

displayed on the screen immediately after the response to the last trial in a block.

## Results

### Alphabet recitation

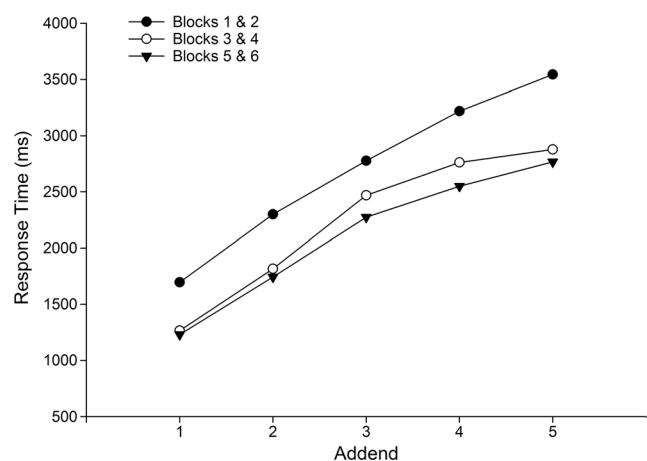
Mean time for the final (i.e., third) alphabet recitation trial was 5.5 s ( $SE = 0.26$ ), similar to the mean alphabet recitation time for the English-first group in Experiment 1 (5.1 s).

### Practice phase

Median RT for correct responses received an ANOVA with practice block as a repeated-measures factor. Mean RT sped up by 30.9% across the six practice blocks with means of 2,926 ms, 2,471 ms, 2,365 ms, 2,215 ms, 2,185 ms, and 2,022 ms,  $F(2.8, 133.67) = 26.54$ ,  $MSE = 316,767$ ,  $p < .0001$ ,  $\eta_p^2 = .36$ . The error rate was 18.3% in Block 1, but errors were less frequent in Blocks 2 to 6 (12.1%, 12.3%, 9.2%, 8.1%, 10.2%),  $F(5, 235) = 6.38$ ,  $p < .0001$ ,  $MSE = 99.23$ ,  $\eta_p^2 = .12$ , for the main effect of block.

With respect to effects of addend size during practice, averaging across all six blocks, mean RT increased as addend size increased from +1 to +5 (1,295 ms, 1,860 ms, 2,422 ms, 2,829 ms, 2,980 ms), presenting both linear,  $F(1, 47) = 203.44$ ,  $p < .0001$ ,  $MSE = 444,173$ ,  $\eta_p^2 = .81$ , and quadratic,  $F(1, 47) = 20.48$ ,  $p < .0001$ ,  $MSE = 161,961$ ,  $\eta_p^2 = .30$ , components, the latter reflecting a deviation below linearity owing primarily to +5. As in Experiment 1, to determine if this RT pattern was present throughout practice, or emerged across practice blocks, we calculated the median RT for each addend combining Blocks 1 and 2, Blocks 3 and 4, and Blocks 5 and 6.

As Fig. 3 shows, early in practice, RT increased linearly with addend size,  $F(1, 47) = 102.22$ ,  $p < .0001$ ,  $MSE =$



**Fig. 3** Mean response time in the practice phase of Experiment 2 by practice blocks and addend



1,001,282,  $\eta_p^2 = .69$ , and there were no significant deviations from linearity (i.e., none of the higher order contrasts were significant; all  $ps > .07$ ). This linear pattern is consistent with exclusive use of a counting-based strategy with each increment requiring a constant amount of time (about 460 ms). In Blocks 3 and 4, there remained a strong linear effect,  $F(1, 47) = 165.66$ ,  $p < .0001$ ,  $MSE = 504,765$ ,  $\eta_p^2 = .78$ , but a statistically clear quadratic component also emerged,  $F(1, 47) = 13.25$ ,  $p = .001$ ,  $MSE = 391,409$ ,  $\eta_p^2 = .22$ , owing primarily to an inflection at +4 with +5 deviating below the linear positions predicted by +1 to +4. The pattern was the same for the final two practice blocks with both linear,  $F(1, 47) = 146.44$ ,  $p < .0001$ ,  $MSE = 494,390$ ,  $\eta_p^2 = .76$ , and quadratic,  $F(1, 47) = 6.60$ ,  $p = .01$ ,  $MSE = 374,643$ ,  $\eta_p^2 = .12$ , components and no other significant higher order trends. Thus, +5 problems presented greater speed up across practice blocks relative to other addend sizes.

As in Experiment 1, the percentage of errors during practice in Experiment 2 presented a linear effect of addend size,  $F(1, 47) = 6.77$ ,  $p = .01$ ,  $MSE = 253.77$ ,  $\eta_p^2 = .13$ , with error rates higher for larger addend sizes from +1 to +5 (9.9%, 7.6%, 10.4%, 14.8%, and 15.8%); but the error rate dip at +2, combined with higher error rates for +4 and +5 compared to +1 to +3, yielded a marginally significant cubic contrast,  $F(1, 47) = 4.09$ ,  $p = .05$ ,  $MSE = 81.56$ ,  $\eta_p^2 = .08$ . We do not think the weak cubic contrast (i.e., there were two inflections in the effects of addend size on errors) is theoretically telling and do not consider it further.

### Test phase

Mean median RT for correct responses confirmed a counting-based strategy for test phase problems with RT increasing from 1,372 to 1,793 to 2,346 ms for +1, +2, and +3, respectively. The corresponding error rates were 11.0%, 10.5%, and 14.1% for +1 to +3, respectively.

To assess transfer effects, median RT for correct answers received a Sequence Type (augend-sequence vs. sequence-only)  $\times$  Transfer Type (transfer vs. control)  $\times$  Block (1 vs. 2) mixed-factor ANOVA. Figure 4 presents the mean median RT as a function of sequence type, transfer type, and block. Participants were faster overall in test Block 2 (1,810 ms) compared to Block 1 (1,971 ms),  $F(1, 47) = 9.83$ ,  $p = .003$ ,  $MSE = 264,095$ ,  $\eta_p^2 = .17$ , but this was qualified by the Block  $\times$  Transfer Type interaction,  $F(1, 47) = 37.16$ ,  $p < .0001$ ,  $MSE = 82,311$ ,  $\eta_p^2 = .44$ . As in Experiment 1, this occurred because transfer problems were faster than control problems in Block 1 (300 ms),  $t(47) = 4.39$ ,  $p < .0001$ ,  $SE = 68.38$ , but not in Block 2 (-57 ms),  $t(47) = 0.88$ ,  $p = .38$ ,  $SE = 64.16$ . Again, as in Experiment 1, the transfer of learning in Block 1 disappeared in Block 2 because control problems sped up relative to Block 1 when they were repeated in Block 2 (342 ms, somewhat less than 543 ms speed up for +1 to +3 practice problems),

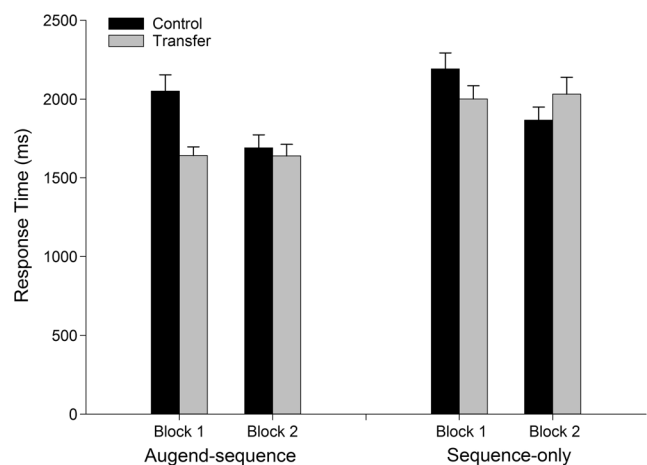
whereas the transfer problems did not show speed up across test blocks (-14 ms; see Fig. 4). The transfer effect in Block 1 occurred both for augend-sequence problems, 411 ms,  $t(47) = 4.15$ ,  $p = .0001$ ,  $SE = 98.96$ , and sequence-only problems, 190 ms,  $t(47) = 2.15$ ,  $p = .037$ ,  $SE = 88.52$ .

Augend-sequence problems were on average 267 ms faster overall than sequence-only problems,  $F(1, 47) = 22.22$ ,  $p < .0001$ ,  $MSE = 308,776$ ,  $\eta_p^2 = .32$ , but there was a Sequence Type  $\times$  Transfer Type interaction,  $F(1, 47) = 4.85$ ,  $p = .03$ ,  $MSE = 236,957$ ,  $\eta_p^2 = .09$ , which occurred because the RT advantage for augend-sequence problems compared to sequence-only problems was larger in the transfer condition (377 ms),  $t(47) = 5.49$ ,  $p < .00001$ ,  $SE = 68.65$ , than in the control condition (158 ms),  $t(47) = 1.94$ ,  $p = .059$ ,  $SE = 81.59$ . Thus, as in Experiment 1, the overall RT advantage for augend-sequence problems owed, in part, to the greater facilitative transfer observed for these items.

Percentage of errors during the test phase also received a Sequence Type  $\times$  Transfer Type  $\times$  Block repeated-measures ANOVA. This revealed fewer errors in Block 2 (9.6%) than in Block 1 (14.1%),  $F(1, 47) = 15.61$ ,  $p = .0003$ ,  $MSE = 125.35$ ,  $\eta_p^2 = .25$ . There also were fewer errors overall for augend-sequence problems (9.4%) than sequence-only problems (14.4%),  $F(1, 47) = 12.67$ ,  $p = .001$ ,  $MSE = 192.02$ ,  $\eta_p^2 = .21$ .

## Discussion

Given that Experiment 2 closely replicated the major results of Experiment 1, we focus here only on the important differences in results. One important difference was that statistically significant transfer occurred both for augend-sequence and sequence-only problems, whereas it occurred only for the former in Experiment 1. Nonetheless, transfer was still greater for augend-



**Fig. 4** Mean response time in the test phase of Experiment 2 as a function of sequence type, transfer type, and block. Error bars are  $\pm 1$  standard error

sequence than sequence-only problems in Experiment 2. This reinforces the conclusion from Experiment 1 that having a common augend in practice and test items enhances transfer of practice in alphabet arithmetic. Another difference in results between experiments was that transfer effects were substantially larger in Experiment 2 than Experiment 1: For augend-sequence problems in Block 1, the transfer effect was 411 ms compared to 282 ms in Experiment 1. For sequence-only problems in Block 1 of Experiment 2, the transfer effect was 190 ms compared to 59 ms in Experiment 1. This could be a result of the interpolated counting task between blocks in Experiment 2, which we introduced to limit use of episodic working memory of the preceding block to solve problems. In theory, this would increase practice of the counting strategy in Experiment 2 relative to Experiment 1, which would be expressed as greater transfer observed in the counting-based performance of new items encountered in the test phase in Experiment 2. Regardless of the cause of the difference, the overall enhanced transfer effect in Experiment 2 made it possible to detect significant generalization of practice for sequence-only transfer items. This is important because it demonstrates that the conditions for generalization of counting practice are quite minimal, requiring only strengthening of access to the relevant counting sequence and does not require practice of a particular starting point (i.e., augend) in the sequence.

### Reanalysis of published addition generalization experiments

Our results showed that augend-sequence matches can contribute to robust generalization of practice in counting-based tasks. Our previous addition generalization experiments (Campbell & Beech, 2014; Chen & Campbell, 2014, 2016) examined generalization between items within the simple-addition problem categories identified by Fayol and Thevenot (2012), but we did not analyze for effects of augend-sequence matches between practiced and tested problems. To pursue this, we conducted new analyses that combined all the data from the three published addition generalization studies. All three experiments were conducted at the University of Saskatchewan, used the same design, procedure and computer-displayed stimuli, but differed in the populations sampled. Campbell and Beech recruited 64 student volunteers from the psychology participant pool, Chen and Campbell (2014) tested 36 Canadian and 36 Chinese adults recruited through the participant pool or by on-line advertisements, and Chen and Campbell (2016, Experiment 2) tested 36 engineering and

computer science students analogous to the engineering students tested by Fayol and Thevenot (2012, Experiment 1). The combined studies totalled 172 participants. The problem sets tested included single-digit plus single-digit addition problems including  $0 + N$ ,  $1 + N$ ,  $N + N$  (ties), other small nonties with sums  $\leq 10$  and large nonties. Each experiment consisted of two blocks of 48 trials. The first randomly selected half of each problem type encountered within each block constituted Subblock A and the second half was Subblock B (for full details of the methods, see Campbell & Beech, 2014; Chen & Campbell, 2014, 2016). Problem order was randomized for each participant; consequently, Subblock A and Subblock B for each problem type (i.e., the first half vs. the second half of each problem type encountered within the block) had random assignment of problems to the two subblocks. In all three of the experiments, only the  $0 + N$  problems presented evidence of generalization of practice across subblocks (i.e., across different problems within each problem type; e.g., practicing  $0 + 3$  facilitated subsequent performance of  $0 + 8$ ).

The reanalysis was particularly focussed on small, nontie additions with sums  $\leq 10$  (including  $1 + N$  problems), which may be the more likely candidates for fast counting-based procedures. The analysis program identified small nontie additions that were preceded earlier in a trial block by one or more small nontie problems that had the same larger addend and the smaller addend was larger (e.g.,  $6 + 1$  preceded by  $6 + 3$  or  $6 + 2$ ). These cases are directly analogous to our alphabet arithmetic stimuli with an augend-sequence match. Maximum search depth was half a block (24 trials), which yielded similar mean numbers of augend-sequence practiced (6.0) and unpracticed (8.4) problems in each of the two trial blocks. Among the 4,980 trials in the analysis, the RTs for 3.4% were marked as spoiled by the experimenter and discarded owing to failures of the microphone to detect the onset of a participant's verbal response. The mean error rate was very low for these problems (0.9% for augend-sequence unpracticed problems and 1.1% for augend-sequence practice problems). For trials with correct answers, in Block 1 the RT means for augend-sequence practiced and unpracticed problems were 736 ms and 736 ms, respectively, and in Block 2 they were 711 ms and 697 ms, respectively. A  $2 \times 2$  ANOVA indicated no main effect of augend-sequence practice or an interaction (both  $ps > .11$  and  $\eta_p^2 \leq .015$ ), but there was a main effect of block with mean RT 33 ms faster in Block 2 when the identical problems were repeated,  $F(1, 171) = 44.49$ ,  $p < .00001$ ,  $MSE = 4072$ ,  $\eta_p^2 = .21$ . Thus, there was no evidence that having recently practiced one or more problems

with a matching augend and counting sequence facilitated addition performance, but there was robust speed up when the identical problem was repeated. This would be expected if performance of the small nontie additions was based on item-specific fact retrieval.<sup>2</sup>

## General discussion

These experiments sought evidence that practice of counting-based procedures can produce generalization of learning to new unpracticed items. This issue is relevant to recent claims that skilled adults' simple addition may be based on fast, automatic counting procedures (e.g., Barrouillet & Thevenot, 2013; Fayol & Thevenot, 2012; Uittenhove et al., 2016) rather than on direct fact retrieval. These experiments demonstrated strong generalization when counting was surely involved; therefore, if a counting process is involved in solving simple-addition facts, there should be comparable, observable transfer, but several experiments have failed to observe such generalization. Campbell and Beech (2014; see also Campbell, Dufour, & Chen, 2014; Chen & Campbell, 2014) showed that practicing the addition identity rule ( $N + 0 = N$ ) on one set of problems facilitated performance for other, unpracticed identity-rule problems, but nonzero addition problems presented no evidence of generalization. The absence of generalization for nonzero additions potentially is evidence against the automatic counting theory, but there was no direct evidence that practice of counting-based procedures should be expected to generalize. In fact, previous research examining alphabet arithmetic problems such as  $B + 3 = E$  found little evidence of transfer of learning to new problems (Logan & Klapp, 1991).

Logan and Klapp (1991) were interested in the development of automaticity of alphabet arithmetic and measured transfer after participants had likely transitioned from counting to memory retrieval to solve individual problems. This potentially made their experiment relatively insensitive to possible generalization of counting practice to new problems. In contrast, we designed the present alphabet arithmetic experiments specifically to measure generalization of counting practice and measured transfer when participants were still relying primarily on counting to solve practice items. Under these conditions, robust generalization of practice was observed in both experiments when practiced problems and new unpracticed test problems shared a common augend letter and the test problems' answer was within the letter-counting

sequence of practiced problems. In Experiment 2, a significant generalization effect also was observed when only a test problems' answer was in one of the letter-counting sequences practiced, thereby demonstrating that a common augend is not necessary to observe transfer from counting practice.

Nonetheless, transfer effects in mean RT for augend-sequence problems were larger relative to sequence-only problems in both experiments. This is consistent with the view that generalization of practice to new test items increases as the number of common elements or processes shared by practice and test items increases (Singley & Anderson, 1989). Our results reinforce the conclusion that counting-based procedures would be expected to yield generalization and reinforces the conclusion that an absence of generalization in adults' nonzero addition (e.g., Chen & Campbell, 2014) represents a genuine challenge to the view that skilled adults' simple addition is based on counting.

Furthermore, our results reinforced the conclusion of Logan and Klapp (1991) that practicing a counting-based procedure leads to memorization of the individual items, rather than automatization of the counting procedure. Although we expected relatively little memorization of the 10 practice items in only six blocks of training, the stimulus set of practice items had a built-in memory-related bias (a manipulation of associative "fan") to favor memorization of +4 and +5 problems because their augend letters were associated with fewer different problems (two each) than +1, +2, or +3 problems (three each). In both experiments, RT as a function of addend size was strictly linear early in practice (Blocks 1 and 2), but developed a marked quadratic component later in practice owing to +4 and +5 problems' mean RT falling below the linear positions predicted by +1, +2, and +3 problems. This would occur if +4 and +5 problems had begun to transition from counting to fact retrieval sooner than +1 to +3, possibly owing to the lower associative fan for the former.

There could be other reasons for the departure from linearity (e.g., +5 problems could be more distinctive and easier to memorize for reasons other than lower fan; see Logan & Klapp, 1991), but the fact that mean RT for +4 and +5 in both experiments was practically identical late in practice is not consistent with performance based exclusively on counting, which should always yield slower average times for +5 than +4. Therefore, the nonlinear effects of addend size that emerged as practice proceeded are strong evidence that transition to retrieval had begun to occur for +5 and perhaps +4 problems. Therefore, the current results provided new evidence that training of a counting-based procedure on a restricted set of items leads to memorization of individual facts rather than to automatization of the procedure. Logan and Klapp used a true–false verification paradigm to study alphabet arithmetic; thus, the current results extend the evidence that counting-based processes are replaced by fact retrieval to a paradigm that required

<sup>2</sup> Uittenhove et al. (2016) proposed that a compacted counting procedure may only develop for the 12 nontie additions with both addends  $\leq 4$ . Restricting the reanalysis of the three generalization studies to only these items depleted the sample size from 172 to only 65 because of empty cells, greatly reducing the power of the analysis. It is worth mentioning, however, that the results were statistically the same as those reported for all small-nontie problems.

participants to produce the correct answers. The production task may be more relevant in relation to acquisition of skill in addition, because answer production would be a more common context for learning addition than learning to discriminate true and false simple addition equations.

In the practice phase of both experiments, the +1, +2, and +3 problems sped up more or less in parallel across practice blocks (see Figs. 1 and 3). If the counting procedure used was speeding up with practice then we might expect greater speed up as the number of incrementing steps increased. Given the structure of our practice sets (e.g.,  $C + 1 = D$ ,  $C + 2 = E$ ,  $C + 3 = F$ ), however, +1 problems are embedded in both +2 problems and +3 problems, and +2 problems are embedded in +3 problems. As a result, in each practice block, the +3 sequences (e.g., count through CDEF) were practiced once, +2 sequences were practiced twice (e.g., count through CDE in both  $C + 3$  and  $C + 2$ ), and the +1 sequences were effectively practiced thrice (e.g., count through CD in  $C + 3$ ,  $C + 2$ , and  $C + 1$ ). RT benefits owing to such embedding is exactly the augend-sequence generalization effect observed in both experiments in the test phase. It will require further experiments, however, to understand the parallel speed up observed across addend size.

In the test phase of both experiments, control problems sped up substantially relative to Block 1 when they were tested again in Block 2, but transfer problems showed no speed up in Block 2 relative to Block 1. We suggested that for control items in Block 2, participants might have sometimes employed relatively fast episodic recall of items from Block 1, whereas episodic memory for transfer items would encounter interference from the practice of related problems during the training phase so that performance of transfer problems relied more or less entirely on counting in both test blocks. This strategy asymmetry would afford RT gains for control problems relative to transfer problems in Block 2.

Introducing the countdown task between blocks in Experiment 2 was intended to inhibit episodic retrieval strategies in Block 2 of the test phase. The same pattern nonetheless emerged whereby transfer problems presented no RT benefit of item repetition across test blocks, whereas control problems presented substantial speed up owing to item repetition. For augend-sequence transfer problems, repetition gains might not occur because generalization of learning from the practice phase limited further opportunities for RT gains, at least given only a single repetition of an item during the test phase. This explanation, however, does not extend to sequence-only problems, which exhibited much smaller generalization effects, but nonetheless showed little or no benefit of a repetition.

One possibility is that the countdown task in Experiment 2 was inadequate to prevent an episodic memory strategy that favoured control items in Block 2, or perhaps the effect is owed to some other mechanism. For example, for transfer problems but not control problems, highly accessible (i.e.,

practiced) procedural components at test may discourage attempts at direct answer retrieval that would exploit episodic memory or alternatively might discourage associating the presented problem with the answer generated by the procedure. We can only speculate about its cause at this point, but as this unexpected effect (i.e., the Transfer Type  $\times$  Block interaction) was robust in both experiments, it is apparently a salient phenomenon of this paradigm. Although novel and interesting in its own right, it is important to emphasize that it has no direct bearing on the evidence for generalization observed in Block 1 of the test phase.

## Conclusions

Like common counting strategies initially used for simple addition, alphabet arithmetic involves enumeration of successive elements in an ordered list. These commonalities make alphabet arithmetic a potentially useful proxy for memory and learning processes that mediate counting strategies in genuine arithmetic (Zbrodoff & Logan, 2005). These experiments provided in-principle evidence that counting-based procedures can produce generalization of learning to unpracticed items. The significant generalization of practice for sequence-only transfer items in Experiment 2 showed that the conditions for generalization of counting practice are quite minimal, requiring only strengthening of access to the local counting sequence, and do not require practice of counting from a particular starting point (i.e., augend). This suggests that generalization of practice owing to use of a counting strategy for simple addition, if it were used by the majority of participants, could be relatively easy to observe, especially when practice and test problems overlap both with respect to the augend and count sequence. Our reanalysis of the previously published addition generalization experiments (Campbell & Beech, 2014; Chen & Campbell, 2014, 2016; combined  $n = 172$ ), however, found no evidence of facilitation when small nontie problems were preceded by problems with a matching augend and counting sequence. Performance of the small additions is not so fast or automatic that they could not benefit from transfer effects.  $N + 1$  problems, for example, display robust speedup when they are tested a second time (Campbell et al., 2013; Chen & Campbell, 2014, 2015), demonstrating that they clearly benefit from practice. Given this, they should similarly be susceptible to generalization effects if they existed. Thus, there remains no evidence of the generalization of practice that might be expected if counting procedures mediated adults' simple addition.



**Author note** This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Jamie Campbell.

## References

- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.
- Ashcraft, M. H., & Guillaume, M. M. (2009). Mathematical cognition and the problem size effect. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 121–151). Burlington, VT: Academic Press.
- Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth grades. *Journal of Experimental Child Psychology*, *33*, 216–234.
- Bajic, D., & Rickard, T. C. (2009). The temporal dynamics of strategy execution in cognitive skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 113–121.
- Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval: Evidence from number and alphabetic arithmetic in children and adults. *Memory & Cognition*, *26*, 355–368.
- Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, *128*, 35–44.
- Campbell, J. I. D., & Beech, L. C. (2014). No generalization of practice for non-zero simple addition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1766–1771.
- Campbell, J. I. D., Chen, Y., & Maslany, A. J. (2013). Retrieval-induced forgetting of arithmetic facts across cultures. *Journal of Cognitive Psychology*, *25*, 759–773.
- Campbell, J. I. D., Dufour, K. D., & Chen, Y. (2014). Retrieval-induced forgetting of multiplication facts and identity rule. *Memory & Cognition*. doi:10.3758/s13421-014-0483-1
- Campbell, J. I. D., & Theriault, N. H. (2013). Retrieval-induced forgetting of arithmetic facts but not rules. *Journal of Cognitive Psychology*, *25*, 717–724.
- Chen, Y., & Campbell, J. I. (2014). Generalization effects in Canadian and Chinese adults' simple addition. *Canadian Journal of Experimental Psychology*, *68*, 152–157.
- Chen, Y., & Campbell, J. I. D. (2015). Operator and operand preview effects in simple addition and multiplication: A comparison of Canadian and Chinese adults. *Journal of Cognitive Psychology*, *27*, 326–334.
- Chen, Y., & Campbell, J. I. D. (2016). Operator priming and generalization of practice in adults' simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 627–635. doi:10.1037/xlm0000196
- Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition*, *123*, 392–403.
- Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, *79*, 329–343.
- Kole, J. A., & Healy, A. F. (2013). Is retrieval mediated after repeated testing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 462–472.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 179–195.
- Pirolli, P., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 136–153.
- Rickard, T. C., & Bajic, D. (2003). Automatic mediation or absence of mediation? Commentary on Crutcher and Ericsson (2000). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1381–1386. doi:10.1037/0278-7393.29.6.1381
- Rickard, T. C., Lau, J. S., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review*, *15*, 656–661.
- Roussel, J. L., Fayol, M., & Barrouillet, P. (2002). Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology*, *14*, 61–104. doi:10.1080/09541440042000115
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Sohn, M.-H., & Carlson, R. A. (1998). Procedural frameworks for simple procedural skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1052–1067.
- Thevenot, C., Barrouillet, P., Castel, C., & Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: A counting model account. *Cognition*, *146*, 289–303.
- Uittenhove, K., Thevenot, C., & Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition*, *146*, 289–303.
- Zbrodoff, N. J. (1995). Why is  $9 + 7$  harder than  $2 + 3$ ? Strength and interference as explanations of the problem-size effect. *Memory & Cognition*, *23*, 689–700.
- 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, NY: Psychology Press.