

Research Reports

Retrieval Priming in Product Verification: Evidence From Retrieval-Induced ForgettingJosh Neudorf*^a, Yalin Chen^a, Jamie I. D. Campbell^a^[a] Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.**Abstract**

The conditions under which multiplication verification ($3 \times 6 = 12$, true or false?) involves product retrieval and comparison or familiarity-based recognition judgements has not been clearly established. In two experiments examining verification of single-digit multiplication problems, we used Retrieval-Induced Forgetting (RIF), a signature of retrieval use, as an index of product retrieval in multiplication verification. In Experiment 1, 72 adults practiced multiplication either in a production format or in a verification format and then were tested on corresponding addition and control problems. The results showed RIF (i.e., slower answer production for addition problems whose multiplication counterparts had been practiced) in both the production-practice and the verification-practice groups, but RIF was stronger following true than false verification. Experiment 2 tested verification with related-false and unrelated-false products. Related-false equations produced longer RTs than unrelated false equations. Practice of true, related-false and unrelated-false multiplication equations all produced RIF of the addition counterparts but, overall, related-false multiplication equations produced relatively weak RIF. The results indicated that product retrieval mediates multiplication verification even when false answers are weak associative lures and suggest that a retrieve-and-compare process is the default strategy when false answers are at least plausible. We conclude that the presented answer in verification equations act as retrieval-priming stimuli with true equations priming correct answer retrieval and related-false answers interfering with correct answer retrieval.

Keywords: multiplication verification, retrieval priming, retrieval-induced forgetting

Journal of Numerical Cognition, 2018, Vol. 4(3), 572–589, doi:10.5964/jnc.v4i3.156

Received: 2017-11-27. Accepted: 2018-03-15. Published (VoR): 2018-12-21.

*Corresponding author at: Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada. E-mail: jamie.campbell@usask.ca



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The arithmetic verification task ($8 \times 4 = 24$, true or false?) has been widely used and studied since the 1980s (e.g., Stazyk, Ashcraft, & Hamann, 1982) and has continued to be a part of diverse research in recent years including neurophysiological studies (e.g., Avancini, Soltész, & Szűcs, 2015; Domahs et al., 2007; Galfano, Penolazzi, Vervaeck, Angrilli, & Umiltà, 2009; Jasinski & Coch, 2012; Núñez-Peña, Gracia-Bafalluy, & Tubau, 2011; Szűcs & Soltész, 2010), cognitive science (e.g., Desmet et al., 2012; Ghirardelli et al., 2010), and educational psychology (e.g., Rotem & Henik, 2013, 2015; van der Ven, Straatemeier, Jansen, Klinkenberg, & van der Maas, 2015). The verification task, and the arithmetic production task in which participants verbally produce an answer, have been mainstays of the cognitive arithmetic research literature for decades (Ashcraft, 1992; Zbrodoff & Logan, 2005). Verification problems may be solved by recognition (i.e., how much familiarity or “resonance” an equation activates; Zbrodoff & Logan, 1990), plausibility-based strategies (e.g., $7 \times 3 = 24$ must be false because two odd numbers yield an odd product; Krueger, 1986; Krueger & Halford, 1984; Lemaire &

Reder, 1999), or verification problems can be solved by a retrieve-and-compare strategy in which the problem's correct answer is generated and directly compared to the presented answer (e.g., Ashcraft, Fierman, & Bartolotta, 1984; Avancini, Soltész, & Szűcs, 2015; Campbell & Fugelsang, 2001; Koshmider & Ashcraft, 1991; Romero, Rickard, & Bourne, 2006).

Despite wide usage of the arithmetic verification task in diverse research, an important unresolved issue concerns what the normative or default strategy is for simple arithmetic verification. Do participants typically produce the correct answer during verification or rely on a familiarity-based recognition strategy or plausibility strategy? When false equations may be consistently identified as false on the basis of a salient characteristic or manipulation (e.g., parity or magnitude disagreement with the correct answer; e.g., Lemaire & Reder, 1999) it is likely that participants do use a familiarity or plausibility check to decide true or false. When participants are not induced to exploit regularities in the experimental stimuli, what is the default strategy? In the present experiments we addressed this question for verification of simple multiplication equations.

There is a long history to this question. With respect to multiplication verification, Campbell (1987) argued that even when false answers make recognition or plausibility judgements difficult or unreliable (e.g., false answers are strong associative lures, such as $8 \times 4 = 24$), verification does not measure the same number-fact retrieval process as the production task. The difference may be explained by priming effects arising from the presented answer to be verified. Meagher and Campbell (1995; see also Campbell, 1987, 1991) measured effects of numerical primes (displayed for 200 ms) on production of multiplication facts (e.g., $4 \times 8 = ?$) with prime-problem inter-stimulus intervals (ISIs) of 0, 750, or 1500 ms. The 0 ISI condition approximates the simultaneous presentation of answer and operands in the standard verification task. Experiment 1 employed three kinds of numerical primes: correct, related, unrelated and a neutral prime (###). Correct primes were the correct answer to the upcoming problem (e.g., 24 for 4×6), a related prime was a multiple of one or the other operands (e.g., 28 or 18), and an unrelated prime was a product of whole number factors but not a multiple of either operand (e.g., 27). Relative to neutral primes, correct primes produced constant RT and accuracy benefits across ISIs, and unrelated primes produced constant RT costs. Related primes produced costs compared to unrelated primes at the 0-ms ISI only. In Experiment 2, eliminating correct-answer primes from the stimulus set eliminated all the false-prime effects except the costs of related primes at the 0-ms ISI.

To explain these findings, Meagher and Campbell (1995) proposed a fast-acting retrieval priming mechanism that yields interference effects for related-false primes and facilitation when the prime is the correct answer. These multiplication retrieval-priming effects were proposed to be automatic consequences of encoding a correct or associatively related numerical prime at the time of product retrieval. If such priming effects are automatic, retrieval-priming effects would also be expected to operate in the multiplication verification task, and similar effects have been observed in the product verification task (Campbell, 1987; Koshmider & Ashcraft, 1991). Priming and interference effects in multiplication production and verification do not provide strong evidence, however, that product retrieval is the default strategy for multiplication verification because they might reflect effects of related primes on equation recognition processes (Zbrodoff & Logan, 1990).

Furthermore, other phenomena raise doubts that product retrieval (e.g., as opposed to recognition of the equation without product retrieval) occurs during multiplication verification. Campbell and Tarling (1996) alternated multiplication production trials (e.g., $9 \times 6 = ?$) with verification trials ($4 \times 9 = 36$, true or false?) and analyzed error priming. Error priming is the phenomenon that simple-arithmetic errors produced in a running sequence of

trials frequently match the answer to a problem solved earlier in the trial block (e.g., 9×4 answered correctly on Trial 10, then observe $9 \times 6 =$ “thirty six” several trials later; Campbell, 1991, 1994; Campbell & Clark, 1989). They found that production errors were strongly primed by previous production trials (the error-answer matching rate was about twice that expected by chance), but production errors were not strongly primed by previous verification trials. Conversely, verification errors were primed by previous verification trials, but not by production trials. Campbell and Tarling (1996) concluded that simple multiplication production and verification were mediated by different memory processes and suggested that a familiarity-based recognition process mediated product verification rather than a retrieve-and-compare process.

The Present Experiments

An alternative approach to investigating retrieval processes in product verification involves retrieval-induced forgetting effects observed in simple addition (e.g., $2 + 3 = ?$) following retrieval practice of the multiplication counterparts (e.g., $2 \times 3 = ?$). Several experiments (e.g., Campbell, Dufour, & Chen, 2015; Campbell & Therriault, 2013; Campbell & Thompson, 2012a) have demonstrated a slowing of RTs and increased errors for addition counterparts following multiplication practice using a variant of the retrieval-practice paradigm developed to study retrieval-induced forgetting (RIF) with verbal materials by Anderson, Bjork, and Bjork (1994). In the original paradigm, several categorical word lists comprising category-cue pairs (e.g., FRUIT-Orange, FRUIT-Banana, PROFESSION-Teacher, PROFESSION-Lawyer) are first viewed, then half of a subset of items in half the categories receive cued-retrieval practice (FRUIT-O, PROFESSION-L) and finally all cues are tested. The typical finding is that unpracticed items from the practiced category are more difficult to remember than items in the unpracticed categories. RIF may reflect inhibition of associative competitors during retrieval practice or cue-based interference (see Storm & Levy, 2012, for a review). In the present context, the critical feature of RIF is the principle that RIF is retrieval dependent: RIF is observed when the practice phase requires fact retrieval but not when the items are studied but not retrieved (see Anderson, 2003; Storm & Levy, 2012). Multiplication-induced RIF of addition counterparts has been repeatedly shown to be retrieval dependent, observed only when multiplication fact retrieval is practiced (e.g., $4 \times 6 = ?$) but not when multiplication equations (e.g., $4 \times 6 = 24$) are studied, even when study practice is as effective as product-retrieval practice at facilitating performance on a subsequent multiplication retrieval test of the practiced facts (Campbell et al., 2013; Campbell & Thompson, 2012a). Galfano, Penolazzi, Fardo, et al. (2011) found that 40 repetitions of passive practice (i.e., study) of simple multiplication equations did result in slower mean RT to verify related multiplication equations. In contrast, our paradigm, which provides only six practice repetitions and tests addition production rather than verification has never found addition RIF with study practice of multiplication facts (Campbell et al., 2013; Campbell & Thompson, 2012a; Maslany & Campbell, 2013).

The retrieval-dependence of multiplication-fact retrieval induced addition RIF provides a diagnostic test for the occurrence of answer retrieval during multiplication verification trials. If product verification entails answer retrieval, rather than only familiarity-based recognition, for example, then subsequent RIF of addition counterparts should be as robust following verification practice as following multiplication production practice. Furthermore, we may expect multiplication RIF to be more robust following true-verification practice than false-verification practice owing to retrieval priming. According to Meagher and Campbell (1995), presented answers should prime correct-product retrieval for true primes but interfere with correct-product retrieval when presented answers are related-false products.

In Experiment 1, two groups of participants received either six practice blocks of product verification trials (e.g., $4 \times 6 = 28$, true or false?) or product production trials (e.g., 4×6 , state the product). False equations involved categorically-related answers: specifically, the correct answer if one operand is changed by ± 1 (e.g., 28 is an exemplar in the factor category 4). The multiplication practice phase was followed by an addition test phase with two addition-production blocks including addition counterparts of practiced multiplication problems and counterpart-unpracticed controls. The addition problems all had sums ≤ 10 or were so-called “tie” problems ($2 + 2$, $3 + 3$, etc.), because larger, non-tie additions (e.g., $9 + 6$) usually do not produce the RIF effect (Campbell & Thompson, 2012a; Campbell et al., 2013; but see Campbell & Dowd, 2012). North American adults often display weak memory strength for the large non-tie additions, making them weaker competitors for their multiplication counterparts and therefore less susceptible to RIF (see Anderson, 2003, for a discussion of competition-dependence and RIF). For the product verification practice group, across the six practice blocks the same multiplication problems appeared with true or false answers in all blocks (e.g., if counterbalancing assigned 4×6 to the true condition, it appeared as $4 \times 6 = 24$ in all six verification blocks). This design feature afforded tests of possible differences between RIF induced by practicing true and practicing related-false verification equations.

Experiment 1

Method

Participants

Seventy-two participants were recruited at the University of Saskatchewan and received course credit or \$7.50. Participants were assigned alternately to the multiplication-production practice or multiplication-verification practice condition yielding two groups of 36. Recruitment materials stipulated English as the first language for elementary arithmetic because addition RIF is potentially sensitivity to linguistic or cultural factors. The effect is robust in English speakers (Campbell & Thompson, 2012a), but Campbell et al. (2013) and Chen and Campbell (2017) did not observe RIF for small non-tie addition problems among Chinese adult participants. The present sample included 50 women and 22 men (65 right-handed, 6 left-handed, and 1 ambidextrous) with mean age of 26.9 years ($SE = 1.06$).

Apparatus

Stimuli were presented using E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2012) on an LED monitor viewed by an experimenter and a CRT monitor viewed by the participant. Black characters in Courier New size 14 font appeared against a white background. The participant sat approximately 50 cm from the monitor, with a hand-held microphone that detected the participant’s voice response and activated a switch that provided the stop signal to a software clock to measure RT.

Stimuli and Design

Participants received six practice blocks of product-verification trials (e.g., $4 \times 6 = 28$, true or false?) or six practice blocks of product-production trials (e.g., $4 \times 6 = ?$, state the product) followed by a two-block addition production test phase (e.g., $4 + 6 = ?$). The multiplication practice phase included primary problems and filler problems (explained further on), and the test phase included only primary problems. For both the verification and production practice groups in both the practice and test phases, problem order was independently randomized for each block.

The primary multiplication and addition stimuli were composed from two sets of numerically small non-tie (sum ≤ 10) and tie (i.e., repeated) operand pairs. For example, the number pair 2 and 5 yielded 2×5 and $2 + 5$ (or the complements) and the pair 44 yielded 4×4 and $4 + 4$. Direct memory retrieval is the predominant strategy reported by educated adults for the small non-tie and tie multiplication and addition problems (Campbell & Alberts, 2009; Campbell & Xue, 2001; LeFevre, Bisanz, et al., 1996). Half of the participants in each of the verification and production practice groups received non-ties with the smaller operand on the left (2×5) and for the other half it was on the right (5×2). Each set of operand pairs was comprised of two subsets, which were used to counterbalance problems across conditions. Set 1 included the pairs 25 28 36 44 77 (Subset 1) and 34 35 26 22 99 (Subset 2). Set 2 included pairs 23 27 46 55 66 (Subset 1) and 24 45 37 33 88 (Subset 2).

For the true-false verification-practice group, assignment of Set 1 and 2 to the multiplication-practiced and multiplication-unpracticed condition, and assignment of the problem subsets to the true-verification and false-verification practice conditions, were fully counterbalanced across participants. Each operand pair appeared consistently as a true or false equation throughout verification practice. For each false verification trial a related-false answer was assigned pseudo-randomly by either increasing or decreasing one operand by one and multiplying it by the other operand (e.g., $4 \times 8 = 24$). False answers were restricted to not equal one of the problem's operands or a multiple of five when 5 was one of the operands.

For the production-practice group, exactly the same counterbalancing procedure with the problem sets and subsets was applied. For this group, who viewed a problem to be answered verbally (e.g., $2 \times 5 = ?$), rather than an equation to be verified (e.g., $2 \times 5 = 10$ or $2 \times 5 = 8$), the counterbalanced assignment of problem subsets to nominal true and false conditions allowed us to treat true vs. false as a properly counterbalanced factor (with respect to problem subsets) in analyses that combined the production and verification groups.

Additionally, during the practice phase both groups also received 10 large (sum > 10) non-tie multiplication problems including 2×9 , 3×8 , 3×9 , 4×7 , 4×8 , 4×9 , 5×6 , 5×7 , 5×8 and 6×7 . These served as filler problems to interfere with verification participants noticing that the same small/tie problems were consistently true or false. In each block, five of the foil problems were randomly selected to be true problems and the other five were false problems.

Following the practice phase, all participants received two test blocks of 20 addition production problems made from all operand pairs in both Sets 1 and Set 2. Addition counterparts of the multiplication-practiced pairs (e.g., $2 + 5$ is the addition counterpart of 2×5) were the RIF targets and the addition counterparts of multiplication-unpracticed pairs served as the control additions.

Procedure

Participants were tested individually in a half-hour session that included a warm-up task preceding the main experimental task. Instructions encouraged both speed and accuracy. In the warm-up, the participant named the eight letters "a" through "h" appearing individually in a random order at the center of the screen. On each trial, a fixation dot appeared at the center of the screen and then flashed twice over a 1-sec interval. On what would have been the third flash, the letter to be named appeared on the screen at fixation. For experimental trials, the fixation dot display was the same as in the warm up task and the problem appeared with the operator (\times or $+$) at fixation. Verification participants were instructed to verbally respond "true" or "false" for verification

trials and production participants were asked to state the correct product. In the addition production test phase, all participants were instructed to state the correct sum. Response timing began when the problem appeared and stopped when the participant's verbal response triggered the voice-activated relay. The spoken response caused the problem to immediately disappear from the screen, which allowed the experimenter to detect and record spoiled RTs where the microphone had failed to detect response onset. After the experimenter entered the given answer or pressed the enter key, the fixation dot for the next trial appeared. There was no feedback about speed or accuracy.

Results

For ANOVA tests, Greenhouse-Geisser corrected statistics were reported when Mauchly's Test indicated violation of the sphericity assumption. Along with null hypothesis significance tests we also reported a Bayes Factor (BF) for each test, calculated using MorePower 6.0 (Campbell & Thompson, 2012b). This program implements the Bayesian Information Criterion (BIC) as proposed by Masson (2011; see also Wagenmakers, 2007), which approximates the unit-information prior as a default, objective Bayes prior probability (Wagenmakers, 2007). The BIC formulation favours H_0 for small effect sizes making it conservative with respect to Type I errors (Nathoo & Masson, 2016). The estimated BF reported is the odds ratio of the null (H_0) over alternative hypothesis (H_1). For example, BF equal to 10 for a given ANOVA test indicates that the data favor H_0 over H_1 by 10 to 1, whereas a value of 0.1 indicates a 10 to 1 ratio in favour of H_1 .ⁱ

Multiplication Practice Phase

RT — A total of 451 practice RTs (5.2%) were marked for exclusion by the experimenter because the voice-key failed to detect response onset, or were discarded as outliers more than 2.5 SD from each Block (1 to 6) × Problem Type (true practiced, false practiced) mean for each participant. The overall error rate (excluding foil problems) during multiplication practice was 3.4%. Mean RT for correct responses received a Practice task (verification vs. production) × Problem type (true vs. false) × Block (1 to 6) ANOVA with practice task as a between-participants factor and problem type and block as repeated-measures factors. True vs. false problem type was a pseudo-factor for the production-practice group. The corresponding means and SEs appear in Table 1.ⁱⁱ

Mean RT followed a decelerating speed-up function across practice blocks (means of 1110, 992, 994, 953, 947 and 943 ms) with greater RT gains from Block 1 to Block 2 than across later practice blocks [$F(3.73, 261.27) = 20.86, p < .005, MSE = 38255, \eta_p^2 = .23, BF < .0001$]. There were no significant interactions involving the block factor (all $ps > .23, \eta_p^2 < .01, BFs > 75000$). Overall, false problems were answered slower than true problems [$F(1, 70) = 32.62, p < .001, MSE = 68124, \eta_p^2 = .32, BF < .0001$], but practice task interacted with problem type [$F(1, 70) = 29.06, p < .001, MSE = 68124, \eta_p^2 = .29, BF < .0001$]. For the verification-practice group, mean RT for false equations (1125 ms) was slower than for true equations (928 ms), whereas for the production-practice group, for which true-false was a pseudo-factor, the nominally true and false problems had practically identical mean RTs (948 ms and 954 ms, respectively).

Table 1

Mean Response Time (SE) and Percentage of Errors by Practice Task, Problem Type, and Block in the Practice Phase of Experiment 1

| Block | Verification Group | | Production Group | |
|----------|--------------------|-----------|------------------|-----------|
| | True | False | "True" | "False" |
| RT | | | | |
| 1 | 1081 (50) | 1229 (60) | 1077 (54) | 1054 (51) |
| 2 | 931 (35) | 1148 (42) | 954 (53) | 936 (48) |
| 3 | 896 (35) | 1146 (49) | 963 (53) | 969 (45) |
| 4 | 884 (39) | 1090 (48) | 896 (45) | 944 (47) |
| 5 | 893 (40) | 1070 (44) | 915 (45) | 912 (45) |
| 6 | 882 (38) | 1067 (49) | 881 (39) | 906 (36) |
| % Errors | | | | |
| 1 | 2.2 (1.1) | 8.3 (2.2) | 2.2 (1.1) | 4.4 (2.1) |
| 2 | 1.1 (0.8) | 6.7 (1.6) | 2.8 (1.8) | 3.3 (1.3) |
| 3 | 1.7 (0.9) | 5.6 (2.1) | 1.1 (0.8) | 3.9 (1.8) |
| 4 | 1.7 (0.9) | 6.1 (1.9) | 1.7 (0.9) | 1.7 (0.9) |
| 5 | 2.2 (1.1) | 6.1 (2.4) | 0.6 (0.6) | 2.8 (1.2) |
| 6 | 1.7 (0.9) | 7.8 (1.8) | 2.8 (1.4) | 2.8 (1.4) |

Note. True vs. false was a pseudo-factor for the production-practice group in that the counterbalancing of problem subsets was the same as that used for the true/false factor for the verification-practice group.

Error rate — Table 1 includes the mean percentage of errors during the practice phase as a function of practice task (verification vs. production), problem type (true vs. false) and block (1 to 6). The corresponding ANOVA indicated a main effect of problem type [$F(1, 70) = 14.16, p < .001, MSE = 151.22, \eta_p^2 = .16, BF = .01$], and there was weak evidence for the same Practice task \times Problem type interaction observed in RT [$F(1, 70) = 4.90, p = .030, MSE = 151.21, \eta_p^2 = .07, BF = .74$], with the verification group making more errors on false equations (6.8%) than true equations (1.8%), whereas the production group had more similar error rates for the nominally false (3.1%) and true (1.9%) problems. There were no other significant omnibus effects (all $ps > .47, \eta_p^2 < .02, BFs > 247000$).

Addition Test Phase

RT — A total of 77 test-phase RTs (2.7% of trials) were marked for exclusion by the experimenter or discarded as outliers more than 2.5 SD from each Block (1, 2) \times Problem Type (true practiced, false practiced, or unpracticed) mean for each participant. The overall error rate during the addition test phase was 2.1% (61 errors). Mean RT for correct responses received a Practice task (verification vs. production) \times Problem type (true practiced, false practiced or unpracticed) \times Block (1 vs. 2) ANOVA with practice task as a between-participants factor and problem type and block as repeated-measures.ⁱⁱⁱ

Mean RT (see Table 2) was faster in Block 2 ($M = 852$ ms, $SE = 21$ ms) than in Block 1 ($M = 921$ ms, $SE = 24$) [$F(1, 70) = 38.23, p < .001, MSE = 13219, \eta_p^2 = .35, BF < .0001$].⁴ RT differed across problem types with means of 905 ms, 897 ms and 856 ms for the true practice, false practice and unpracticed conditions respectively [$F(2, 140) = 8.59, p < .001, MSE = 11656, \eta_p^2 = .11$ for the omnibus test, $BF = .04$], with weak evidence for a linear component of the three-way interaction [$F(1, 70) = 4.55, p = .036, MSE = 8182, \eta_p^2 = .06, BF = .88$].

Table 2

Mean Response Time (SE) and Mean Percentage of Errors (SE) by Practice Group, Problem Type, and Block in the Test Phase of Experiment 1.

| Block | Verification-Practice Group | | | Production-Practice Group | | |
|-----------------|-----------------------------|---------------|----------------|---------------------------|---------------|----------------|
| | Unpracticed | Δ True | Δ False | Unpracticed | Δ True | Δ False |
| RT | | | | | | |
| 1 | 841 (30) | 85 (21)*** | 50 (15)** | 930 (41) | 31 (25) | 44 (33) |
| 2 | 801 (30) | 21 (20) | 3 (18) | 853 (28) | 59 (22)* | 67 (21)** |
| <i>M</i> | 821 (29) | 53 (14)*** | 27 (12)* | 892 (33) | 45 (18)* | 56 (22)* |
| % Errors | | | | | | |
| 1 | 2.2 (0.8) | -1.7 (0.9) | 0.6 (1.5) | 1.4 (0.6) | 3.1 (1.6) | 1.9 (1.7) |
| 2 | 1.4 (0.6) | -0.8 (0.6) | 1.9 (1.7) | 1.4 (0.7) | 1.9 (1.4) | 1.4 (1.2) |
| <i>M</i> | 1.8 (0.5) | -1.3 (0.5) | 1.3 (1.3) | 1.4 (0.4) | 2.5 (1.2) | 1.7 (1.2) |

Note. Δ True and Δ False represent potential RIF effects (i.e., difference RT relative to the unpracticed condition as subtrahend) for true and false problems respectively. True vs. false was a pseudo-factor for the production-practice group in that the counterbalancing of problem subsets was the same as for the verification-practice group.

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$. With $df = 35$.

To pursue this, we computed for each participant the mean difference between the true practiced and unpracticed (i.e., baseline) condition, and between the false practiced and unpracticed (i.e., baseline) condition. These two difference scores represent potential RIF effects generated by true and false practice trials, respectively. Positive differences correspond to longer RT in the practiced condition (i.e., RIF). A Block \times Problem Type ANOVA was conducted for each group (i.e., verification or production practice task). The true vs. false factor is a pseudo-variable for the production practice group. The corresponding means and SEs appear in Table 2.

The analysis of the verification-group data provided weak evidence for a larger RIF effect in Block 1 (68 ms, $SE = 15.3$) than Block 2 (12 ms, $SE = 17.0$) [$F(1, 35) = 5.45$, $p = .025$, $MSE = 20047$, $\eta_p^2 = .14$, $BF = .44$], which has been a common finding in arithmetic RIF (Campbell et al., 2013; Campbell & Thompson, 2012a). There was also weak evidence for larger RIF for true-practiced problems (53 ms, $SE = 13.6$) than false-practiced problems (27 ms, $SE = 11.7$) [$F(1, 35) = 4.68$, $p = .037$, $MSE = 5359$, $\eta_p^2 = .12$, $BF = .63$]. The corresponding analysis of the production-task group indicated no significant effects of block or problem type (all $ps > .38$), but both groups presented evidence of RIF overall: 50.3 ms for the production group [$t(35) = 3.40$, $p = .002$, $SE = 14.8$, $\eta_p^2 = .25$, $BF = .03$] and 40.0 ms for the verification group [$t(35) = 3.60$, $p = .001$, $SE = 11.1$, $\eta_p^2 = .27$, $BF = .02$]. Thus, the current experiment provided strong evidence that both the multiplication production and multiplication verification tasks induced RIF of the addition counterparts expressed in verbal production RTs.

Error rate — Table 2 includes the mean percentage of test phase errors by practice task (verification or production), problem type (true practiced, false practiced, or unpracticed) and block (1 and 2). There was a total of 61 errors in the test phase, 2.1% of trials. Of the 432 Participant \times Problem type \times Block cells, 375 (86.8%) contained zero errors. The preponderance of cells at the measurement floor precluded detailed inferential analyses of test phase errors.

Discussion

The finding of robust addition RIF from practicing multiplication counterparts in a verification task implies that the multiplication verification problems were solved using a retrieve-and-compare strategy (e.g., Koshmider & Ashcraft, 1991; Romero et al., 2006) rather than solved by evaluating equation familiarity (i.e., recognition) without explicit retrieval of the correct product (e.g., Campbell & Tarling, 1996; Zbrodoff & Logan, 1990). This conclusion follows because the addition RIF effect has been repeatedly demonstrated to be retrieval dependent, and not observed when the multiplication equations are only studied (Campbell et al., 2013; Campbell & Thompson, 2012a). The overall RIF effect size on addition RT was about the same following verification practice (40 ms) and production practice (50 ms). There was weak evidence that RIF for the verification practice group was greater in connection with true (53 ms) than false (27 ms) verification trials, although the RIF effect was significant at .05 for both types. This fits with the proposal by Campbell (1987) that true verification equations are more likely to prime retrieval of the problem's correct answer than are false verification trials. Campbell (1987) argued that related-false verification products actively interfered with retrieval of the correct answer and promoted retrieval errors. This would contribute to slower RT and higher error rate for related-false than true equations and also contribute to weak addition RIF from related-false verification equations.

Experiment 1 provided evidence that verification of simple multiplication equations was solved by a retrieve-and-compare strategy, but is this finding owed to using closely related false answers? A plausible consequence of using related-false products is that discrimination of true and false equations based only on familiarity was not a viable strategy because both trial types would produce a strong familiarity response (e.g., $2 \times 8 = 14$ might seem initially plausible). This could discourage use of familiarity information to perform the verification task and promote predominant use of a retrieve-and-compare strategy. Experiment 2 pursued RIF of addition fact retrieval manipulating the relatedness of false verification answers during the multiplication practice phase. One group constituted a replication of the verification condition in Experiment 1 in which false multiplication answers were categorically related (i.e., a multiple of one of the factors and the correct answer if one operand was changed by ± 1 ; e.g., $2 \times 8 = 14$). For the second group in Experiment 2, the false answers were unrelated in that they were not a multiple of either operand; e.g., $2 \times 8 = 21$). Participants find it much easier to reject such unrelated-false answers compared to related-false answers (Campbell, 1987; Koshmider & Ashcraft, 1991). Having all false equations appear with an unrelated-false product (e.g., $2 \times 8 = 21$) could increase the utility of a familiarity-based recognition strategy because true equations (e.g., $2 \times 8 = 16$) and unrelated-false equations (e.g., $2 \times 8 = 21$) may be readily discriminable based on familiarity. In this case, in Experiment 2 we would expect to observe RIF for the related-false multiplication group as in Experiment 1 but not the unrelated-false multiplication group because using a recognition-based strategy that does not require explicit retrieval of the correct product should not produce RIF of addition counterparts.

Experiment 2

Method

Seventy-two participant who did not participant in Experiment 1 were recruited in the same way as in Experiment 1. The sample included 51 women and 21 men with a mean age of 22.7 years ($SE = 0.60$). There were 66 right-handed, 5 left-handed, and 1 ambidextrous participant. Experiment 2 was the same as Experiment 1

except that the production group was replaced with an unrelated-false verification group. Unrelated false answers were created by randomly selecting per trial one of the four answers produced by adding or subtracting 1 to/from both operands and then using the product as the false answer. In the case that no non-multiples were produced with this method, the false answer was the true answer plus or minus 1.

Results

Multiplication Practice Phase

RT — A total of 199 practice RTs (2.3%) were excluded as in Experiment 1. The overall error rate during multiplication practice was 4.5%. Mean RT for correct responses received a False-answer type (related practice group vs. unrelated practice group) \times Problem type (true vs. false) \times Block (1 to 6) ANOVA with false-answer type as a between-participants factor and problem type and block as repeated-measures factors. The corresponding mean RTs and SEs appear in Table 3.

Table 3

Mean Response Time (SE) and Mean Percentage of Errors (SE) by Practice Task, Problem Type, and Block in The Practice Phase of Experiment 2.

| Block | Unrelated False Group | | Related False Group | |
|----------|-----------------------|-----------|---------------------|-----------|
| | True | False | True | False |
| RT | | | | |
| 1 | 958 (44) | 1079 (42) | 940 (33) | 1191 (57) |
| 2 | 838 (39) | 987 (44) | 873 (36) | 1044 (44) |
| 3 | 797 (38) | 986 (57) | 822 (32) | 1042 (40) |
| 4 | 814 (34) | 945 (43) | 782 (31) | 1006 (49) |
| 5 | 820 (38) | 932 (41) | 794 (28) | 950 (37) |
| 6 | 784 (30) | 896 (40) | 783 (30) | 965 (40) |
| % Errors | | | | |
| 1 | 5.6 (1.7) | 7.8 (2.0) | 3.3 (1.3) | 7.8 (2.7) |
| 2 | 0.6 (0.6) | 5.6 (2.2) | 1.7 (0.9) | 7.2 (2.0) |
| 3 | 2.8 (1.4) | 5.0 (1.9) | 2.2 (1.1) | 8.9 (2.8) |
| 4 | 2.8 (1.2) | 3.9 (1.6) | 3.3 (1.5) | 7.8 (2.9) |
| 5 | 2.8 (1.4) | 3.3 (1.5) | 1.1 (0.8) | 5.6 (2.2) |
| 6 | 1.7 (0.9) | 5.6 (1.9) | 3.9 (1.8) | 7.2 (2.3) |

As in Experiment 1, mean RT followed a decelerating speed-up function across practice blocks with means of 1042, 935, 912, 887, 873, and 857 ms [$F(3.642, 254.955) = 36.82, p < .001, MSE = 24099, \eta_p^2 = .35, BF < .0001$]. There were no significant interactions involving the block factor (all $p > .05$). False equations were answered slower than true equations [$F(1, 70) = 195.32, p < .001, MSE = 31311, \eta_p^2 = .74, BF < .0001$], but the between group false-answer type factor interacted with problem type [$F(1, 70) = 7.28, p = .009, MSE = 31311, \eta_p^2 = .09, BF = .24$]: The two groups had similar mean RTs for true equations (835 ms and 832 ms for the unrelated-false and related-false groups, respectively), but related-false equations were answered slower on average (1033 ms) than unrelated-false equations (971 ms). There were no other significant omnibus effects (all $p > .06$).

Error rate — Table 3 presents mean error rates during the practice phase as a function of false-answer type (related vs. unrelated), problem type (true vs. false) and block (1 to 6). The only significant test was the main effect of problem type [$F(1, 70) = 10.19, p = .002, MSE = 283.64, \eta_p^2 = .09, BF = .06$; all other p -values $> .06$]. False equations had a higher error rate (6.3%) than true equations (2.6%).

Addition Test Phase

RT — A total of 77 test RTs (2.7%) were discarded as outliers as in Experiment 1. The error rate was 3.3% of trials. Mean RT for correct trials was analyzed as in Experiment 1 and the means and SEs appear in Table 4. We conducted a 2 (False-answer type group: related or unrelated during practice) \times 2 (Block: 1 or 2) \times 3 (Problem Type: true practiced, false practiced or unpracticed) ANOVA with false-answer type a between-participants measure and block and problem type repeated measures. The ANOVA indicated that mean RT was faster in Block 2 (774 ms) than in Block 1 (945 ms) [$F(1, 70) = 53.84, p < .001, MSE = 9996, \eta_p^2 = .44, BF < .0001$]. There was weak evidence that mean RT differed across problem types with means of 822 ms, 816 ms and 790 ms for the true-practiced, false-practiced and unpracticed conditions respectively [$F(2, 140) = 4.185, p = .017, MSE = 10295, \eta_p^2 = .06$, but with $BF = 2.20$, the Bayesian analysis slightly favored H_0]. This was qualified by the linear component of the three-way interaction [$F(1, 70) = 9.11, p = .004, MSE = 4710, \eta_p^2 = .12, BF = .10$].

Table 4

Mean Response Time (SE) and Mean Percentage of Errors (SE) by Practice Group, Problem Type, and Block in the Test Phase of Experiment 2

| Block | Unrelated False Group | | | Related False Group | | |
|----------|-----------------------|---------------|----------------|---------------------|---------------|----------------|
| | Unpracticed | Δ True | Δ False | Unpracticed | Δ True | Δ False |
| RT | | | | | | |
| 1 | 838 (36) | 62 (23)* | 70 (25)** | 794 (27) | 25 (22) | 16 (17) |
| 2 | 795 (28) | -9 (16) | 9 (22) | 732 (25) | 51 (16)** | 13 (14) |
| <i>M</i> | 816 (31) | 26 (16) | 40 (19)* | 763 (25) | 38 (16)* | 14 (11) |
| % Errors | | | | | | |
| 1 | 4.7 (1.4) | 3.1 (2.5) | 0.8 (1.5) | 1.1 (0.5) | < 0.01 (0.8) | 4.4 (2.0) |
| 2 | 4.4 (1.2) | -0.6 (1.4) | -2.2 (1.3) | 1.4 (0.7) | 0.8 (1.1) | 0.3 (1.1) |
| <i>M</i> | 4.6 (1.2) | 1.3 (1.4) | -0.7 (0.9) | 1.3 (0.5) | 0.4 (0.7) | 2.4 (1.3) |

Note. Δ True and Δ False represent potential RIF effects (i.e., difference RT relative to the unpracticed condition as subtrahend) for true and false problems respectively.

* $p \leq .05$. ** $p \leq .01$. With $df = 35$.

As in Experiment 1, to pursue the three-way interaction a Block \times Problem Type ANOVA was conducted for each group (i.e., unrelated-false verification and related-false verification groups) on RT estimates of RIF for each cell (i.e., mean addition RT for true-multiplication practiced minus unpracticed, and mean RT for false-multiplication practiced minus unpracticed). The analysis of the unrelated-false group data indicated a larger RIF effect in Block 1 (66.1 ms) [$t(35) = 3.54, p = .001, SE = 18.7, \eta_p^2 = .26, BF = .02$] than Block 2 (-0.01 ms) [$t(35) = -.001, p = .999, SE = 13.6, \eta_p^2 < .001, BF = 6.00$], with the test for the main effect of block indicating $F(1, 35) = 11.76, p = .002, MSE = 13361, \eta_p^2 = .25, BF = .03$. There was no main effect of problem type or Block \times Problem Type interaction (both $p > .5, BF > 5.0$). Thus, there was no evidence that addition RIF differed as a function of problem type (i.e., true equations vs. unrelated-false equations).

The corresponding analysis of the related-false group indicated no significant effects, although the test for the main effect of problem type (true practice vs. false practiced) was in the same direction as observed in Experiment 1 with RIF of 38 ms for true-practiced problems [$t(35) = 2.33$, $p = .025$, $SE = 16.4$, $\eta_p^2 = .13$, $BF = .45$] compared to 14 ms for related-false practiced problems [$t(35) = 1.27$, $p = .213$, $SE = 11.4$, $\eta_p^2 = .04$, $BF = 2.67$] with $F(1, 35) = 3.13$, $p = .09$, $MSE = 6400$, $\eta_p^2 = .08$, $BF = 1.28$, for the main effect of problem type. There was weak evidence of greater RIF following true than following related-false multiplication verification for this group in Block 2 (51 ms for true vs. 13 ms for false; $t(35) = 2.22$, $p = .033$, $SE = 17.3$, $\eta_p^2 = .12$, $BF = .59$).

Error rate — Table 4 includes the mean percentage of test phase errors by false-answer type, practice type (true practiced, false practiced, or unpracticed) and block (1 and 2) in Experiment 2. Of the 432 Participant \times Practice Type \times Block cells, 353 (81.7%) contained a value of 0. We did not pursue inferential analyses of test phase errors.

Discussion

During the multiplication-practice phase, unrelated-false equations were answered faster on average than related-false equations (971 ms vs. 1033 ms), but this difference between the groups was not observed for true equations (835 ms vs. 832 ms). Thus, as expected, unrelated-false equations were relatively faster to be identified as false compared to related-false equations. Nonetheless, in the test phase, having answered unrelated-false multiplication equations produced quite robust RIF of the addition counterparts in Block 1 (70 ms), suggesting that retrieval of correct products occurred for the unrelated-false multiplication problems in the practice phase. The related-false replication group produced significant RIF only for true equations but the evidence was weak that the addition RIF effect from practice of true multiplication equations was statistically greater than from practice of related-false equations and only observed in test Block 2. When the two related-false verification groups from Experiments 1 and 2 were combined, the observed addition RIF averaged across blocks was approximately twice as large following practice of true product-verification equations (46 ms, $SE = 10.6$) compared to following related-false equations (21 ms, $SE = 8.1$) [$F(1, 70) = 7.76$, $p = .007$, $MSE = 5799$, $\eta_p^2 = .10$, $BF = .20$]. Thus, the combined experiments provided positive evidence that verification of true multiplication equations induced stronger RIF in addition counterparts than did practice of related-false equations.

General Discussion

Arithmetic verification is a widely-used experimental task, but what type of cognitive skill does it measure? Two experiments were designed to use RIF of addition fact retrieval as a diagnostic tool to assess whether multiplication product retrieval or a familiarity-based recognition strategy was used to solve true-false multiplication verification equations. Selection between these two strategies has been assumed to depend on familiarity or plausibility of the false answers used. We proposed that addition RIF from multiplication retrieval practice is indicative of answer retrieval rather than only a familiarity/plausibility check for multiplication verification because addition RIF has repeatedly been demonstrated to be retrieval dependent (e.g., Campbell & Thompson, 2012a; Campbell et al., 2013). Merely studying correct multiplication equations (e.g., $2 \times 3 = 6$) does not produce a subsequent slowing of correct answer retrieval for the addition counterparts ($2 + 3 = ?$), even when study practice is as effective as retrieval practice in facilitating subsequent product retrieval of the studied multiplication problems (Campbell & Thompson, 2012a). Accordingly, we would expect to observe addition RIF following

practice of multiplication verification only if correct product retrieval occurred during verification performance. Thus, the addition RIF effect provides indirect evidence that multiplication answer retrieval occurred during the practice phase.

Experiment 1 used related-false verification products (e.g., $6 \times 4 = 28$), which are products categorically related to one of the problem factors (i.e., operands). These would be expected to be relatively difficult to identify as false based on familiarity alone and therefore likely to promote a retrieve-and-compare strategy. Strong addition RIF was observed from verification practice of multiplication counterparts and it was not statistically different in effect size compared to the RIF produced by answer-production multiplication practice. Nonetheless, there was evidence in Experiment 1 that true-verification produced a larger addition RIF effect on addition counterparts than did false-verification multiplication practice. The related-false replication group in Experiment 2 did not produce as strong evidence as Experiment 1 that true multiplication verification yielded stronger addition RIF than related-false verification, but the effect was present for this group in Block 2 (51 ms for true vs. 13 ms for false), and evidence for the effect was positive when the two related-false verification groups from Experiments 1 and 2 were combined, with RIF averaged across blocks approximately twice as large for true as for related-false equations (46 ms vs. 21 ms). [Campbell \(1987\)](#) proposed that true verification equations primed retrieval of the correct product, whereas related-false equations interfered with retrieval of the correct product and often induced retrieval errors (see also [Meagher & Campbell, 1995](#); [Romero et al., 2006](#), p. 106). It follows that, because addition RIF is retrieval dependent, correct-answer priming and related-answer interference effects could contribute to stronger addition RIF from practice of true than related-false multiplication verification equations.

Experiment 2 introduced an unrelated-false multiplication practice condition. Although unrelated-false verification was easier (i.e., faster) than related-false verification, which might have induced a familiarity-based recognition strategy that would not produce addition RIF, we nonetheless observed addition RIF from practice of unrelated-false multiplication equal in magnitude to true-verification equations. This is consistent with correct-product retrieval mediating both true and unrelated-false verification equations. Perhaps correct answer priming on true trials promotes retrieve-and-compare more generally, at least when false answers are somewhat plausible. The similar RIF effect size for true and unrelated-false multiplication equations suggests that retrieve-and-compare is the default strategy under these experimental conditions. Finding addition RIF from verification practice of true, related-false and unrelated-false trials does not imply that a retrieve-and-compare strategy was used exclusively for multiplication verification in the present studies. Indeed, [Romero et al. \(2006, Experiment 1\)](#) found that North American university students reported using retrieve-and-compare on 72% of multiplication verification equations involving ties (e.g., $6 \times 6 = 36$) and small non-ties with a sum ≤ 10). Consistent with this, our results imply an answer-retrieval-based strategy was used for a sufficiently large proportion of true, related-false and unrelated-false multiplication verification trials to induce a robust RIF effect on the addition counterpart problems.

Reconciling the RIF and Error-Priming Evidence

Arithmetic RIF and error priming in arithmetic both reflect associative competition in retrieval but the two phenomena may arise through different mechanisms. As explained previously, [Campbell and Tarling \(1996\)](#) found in blocks of alternating production and verification trials that error priming was task specific (i.e., correct production trials such as $8 \times 3 =$ “twenty four” primed future production errors such as $8 \times 4 =$ “twenty four”, but did not prime future verification errors, and vice versa). They concluded that multiplication production and verification

were mediated by different memory processes and suggested a familiarity-based over a retrieval-based model of arithmetic verification. In contrast to this same-task dependency for error priming, the present results provided good evidence that both multiplication production and verification involved retrieval of the correct product, as evidenced by RIF of the addition counterparts. This is not the only dissociation observed between error priming and RIF in arithmetic. [Campbell \(1994\)](#) examined error priming in both simple addition and multiplication in separate blocks with the problem format (Arabic digits or English number words) alternating across trials. Both multiplication and addition presented format-specific error priming but no cross-format error priming (i.e., Arabic format problems primed errors on later Arabic-format problems but not on word-format problems, and vice versa). Nonetheless, [Campbell and Thompson \(2012a\)](#) found that practicing multiplication facts in visual number word format (e.g., three \times four) induced RIF measured in RT for addition counterparts tested in digit format (e.g., $3 + 4$).

Thus, although error priming and RIF in number-fact retrieval are both indicators of retrieval competition, they apparently arise from distinct mechanisms. Error priming is sensitive to the conditions of problem encoding (verification equation vs. production problem; Arabic digit vs. number word format) whereas arithmetic RIF is less sensitive to these factors. Error priming has been shown to be stronger when the priming problem and error-primed problem have the same operand in the same left or right position. For example, solving 6×4 is more likely to prime its answer (24) as an error to 8×4 than to 4×8 ([Arbuthnott & Campbell, 1996](#)). Error priming therefore may reflect interference resulting from re-instantiation of a recent retrieval pathway when the current problem shares encoding surface features with a previous retrieval episode (e.g., a common operand, operand surface format and operand spatial position). As noted previously, study of multiplication facts (i.e., viewing and reading aloud $3 \times 4 = 12$) does not induce RIF in addition counterparts even when the study phase produces robust facilitation in a subsequent product production task ($3 \times 4 = ?$; [Campbell & Thompson, 2012a](#)). This indicates that priming of problem surface features and increasing the memory strength of the studied problem is insufficient to produce RIF. Instead, addition RIF effects may reflect inhibition of retrieval competitors during multiplication retrieval practice and this inhibitory mechanism of interference occurs when the target is successfully retrieved regardless of the similarity of practice and test problem format (see [Storm & Levy, 2012](#), for a review of the status of an inhibition theory of RIF).

Conclusion

The present experiments provided strong evidence that multiplication verification produced RIF in addition counterparts expressed in slower addition RTs. Multiplication-induced RIF of addition counterparts has been repeatedly shown not to occur when multiplication equations are studied and no answer is generated, as would normally be the case if multiplication verification was solved by a familiarity-based recognition strategy to discriminate true from false equations. Consequently, the present results are strong evidence that product retrieval occurred in multiplication verification, even for false equations with weakly associated presented answers (i.e., the unrelated false products used in Experiment 2). We propose that except under conditions in which participants are induced to use logical criteria (e.g., odd even agreement of presented and correct answer) or that answers are very implausible (remote in numerical magnitude from correct) that retrieval of the product is normally a routine stage of the verification process ([Ashcraft et al., 1984](#); [Koshmider & Ashcraft, 1991](#)), at least for the small and tie simple multiplication problems that induce RIF in addition counterparts ([Campbell & Thompson, 2012a](#)). The results also suggest that verification practice can be a useful task to facilitate multiplication production learning, given that answer memory retrieval is routinely involved in verification practice.

Notes

- i) Bayes factor is a continuous scale but a conventional interpretation is that a BF greater than 10 or less than 0.1 provides relatively strong evidence for H_0 or H_1 , respectively, whereas BF between 3 and .33 provides little evidence one way or the other for H_0 or H_1 (e.g., Wetzels, van Ravenzwaaij, & Wagenmakers, 2015). Intermediate values of BF from 3 to 10 and .33 to .1 provide graded amounts of evidence for H_0 or H_1 , respectively. The MorePower 6.0 program used to calculate BF in the present article is freely available at <https://wiki.usask.ca/pages/viewpageattachments.action?pageId=420413544>. See Rouder, Morey, Speckman, and Province (2012) and Rouder, Speckman, Sun, Morey, and Iverson (2009) for alternative approaches to Bayes factor calculations for ANOVA and t-tests with online calculations available at <http://pcl.missouri.edu/bayesfactor>. Note that a null hypothesis rejected under standard significance testing assumptions (i.e., $p \leq .05$) can have positive support under Bayesian assumptions (see Masson, 2011, p. 688; Wagenmakers, 2007, p. 793). We focus mainly on the omnibus factorial tests of the practice phase ANOVA because we had no specific, theoretically important hypotheses for the higher-order contrasts involving the block factor.
- ii) We used mean rather than median RT because the true and false conditions and unpracticed condition were not based on the same number of observations (5 and 10 trials respectively). As sample size decreases, the sample median RT increasingly overestimates the population median RT (Miller, 1988).
- iii) In Table 2, absolute RTs are reported only for the unpracticed condition in the test phase. For the true and false practice conditions, we report the mean and SE for mean difference RT relative to the unpracticed condition. The mean difference scores (labelled ΔTrue and ΔFalse) represent potential transfer effects (e.g., RIF) from the practice phase.

Funding

This research was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Jamie Campbell.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

Data Availability Statement

As researchers we did not have research ethics approval at the time the paper was accepted for publication to publish individual participant's data. Please contact the authors for further information.

References

- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49, 415-445. doi:10.1016/j.jml.2003.08.006
- Anderson, M. C., Bjork, E. L., & Bjork, R. A. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063-1087. doi:10.1037/0278-7393.20.5.1063
- Arbuthnott, K. D., & Campbell, J. I. D. (1996). Effects of operand order and problem repetition on error priming in cognitive arithmetic. *Canadian Journal of Experimental Psychology*, 50, 182-195. doi:10.1037/1196-1961.50.2.182
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of theory and data. *Cognition*, 44, 75-106. doi:10.1016/0010-0277(92)90051-I

- Ashcraft, M. H., Fierman, B. A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review, 4*, 157-170. doi:10.1016/0273-2297(84)90005-4
- Avancini, C., Soltész, F., & Szűcs, D. (2015). Separating stages of arithmetic verification: An ERP study with a novel paradigm. *Neuropsychologia, 75*, 322-329. doi:10.1016/j.neuropsychologia.2015.06.016
- Campbell, J. I. D. (1987). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*, 349-364. doi:10.3758/BF03197037
- Campbell, J. I. D. (1991). Conditions of error priming in number-fact retrieval. *Memory & Cognition, 19*, 197-209. doi:10.3758/BF03197119
- Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53*, 1-44. doi:10.1016/0010-0277(94)90075-2
- Campbell, J. I. D., & Alberts, N. A. (2009). Operation-specific effects of numerical surface form on elementary calculation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 999-1011. doi:10.1037/a0015829
- 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. doi:10.1080/20445911.2013.820191
- Campbell, J. I. D., & Clark, J. M. (1989). Time course of error priming in number-fact retrieval: Evidence for excitatory and inhibitory mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 920-929. doi:10.1037/0278-7393.15.5.920
- Campbell, J. I. D., & Dowd, R. (2012). Inter-operation transfer in Chinese-English bilinguals' arithmetic. *Psychonomic Bulletin & Review, 19*, 948-954. doi:10.3758/s13423-012-0277-z
- Campbell, J. I. D., Dufour, K., & Chen, Y. (2015). Retrieval-induced forgetting of multiplication facts and identity rule. *Memory & Cognition, 43*, 672-680. doi:10.3758/s13421-014-0483-1
- Campbell, J. I. D., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition, 80*, B21-B30. doi:10.1016/S0010-0277(01)00115-9
- Campbell, J. I. D., & Tarling, D. P. M. (1996). Retrieval processes in arithmetic production and verification. *Memory & Cognition, 24*, 156-172. doi:10.3758/BF03200878
- Campbell, J. I. D., & Theriault, N. H. (2013). Retrieval-induced forgetting of arithmetic facts but not rules. *Journal of Cognitive Psychology, 25*, 717-724. doi:10.1080/20445911.2013.798328
- Campbell, J. I. D., & Thompson, V. A. (2012a). Retrieval-induced forgetting of arithmetic facts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 118-129. doi:10.1037/a0025056
- Campbell, J. I. D., & Thompson, V. A. (2012b). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods, 44*, 1255-1265. doi:10.3758/s13428-012-0186-0
- Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*(2), 299-315. doi:10.1037/0096-3445.130.2.299
- Chen, Y., & Campbell, J. I. D. (2017). An evaluation of sex and cultural differences in arithmetic retrieval-induced forgetting. *Journal of Cognitive Psychology, 29*(8), 949-962. doi:10.1080/20445911.2017.1348356

- Desmet, C., Imbo, I., Brauwer, J. D., Brass, M., Fias, W., & Notebaert, W. (2012). Error adaptation in mental arithmetic. *Quarterly Journal of Experimental Psychology*, 65(6), 1059-1067. doi:10.1080/17470218.2011.648943
- Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., & Nuerk, H.-C. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions*, 3, Article 66. doi:10.1186/1744-9081-3-66 Retrieved from <http://www.biomedcentral.com/content/pdf/1744-9081-3-66.pdf>
- Galfano, G., Penolazzi, B., Fardo, F., Dhooge, E., Angrilli, A., & Umiltà, C. (2011). Neurophysiological correlates of retrieval-induced forgetting in multiplication fact retrieval. *Psychophysiology*, 48, 1681-1691. doi:10.1111/j.1469-8986.2011.01267.x
- Galfano, G., Penolazzi, B., Vervaeck, I., Angrilli, A., & Umiltà, C. (2009). Event-related brain potentials uncover activation dynamics in the lexicon of multiplication facts. *Cortex*, 45, 1167-1177. doi:10.1016/j.cortex.2008.09.003
- Ghirardelli, T. G., Mills, C. B., Zilioli, M. K. C., Bailey, L. P., & Kretschmar, P. K. (2010). Synesthesia affects verification of simple arithmetic equations. *The Journal of General Psychology*, 137(2), 175-189. doi:10.1080/00221301003645152
- Jasinski, E. C., & Coch, D. (2012). ERPs across arithmetic operations in a delayed answer verification task. *Psychophysiology*, 49(7), 943-958. doi:10.1111/j.1469-8986.2012.01378.x
- Koshmider, J. W., & Ashcraft, M. H. (1991). The development of children's mental multiplication skills. *Journal of Experimental Child Psychology*, 51, 53-89. doi:10.1016/0022-0965(91)90077-6
- Krueger, L. E. (1986). Why $2 \times 2 = 5$ looks so wrong: On the odd-even rule in product verification. *Memory & Cognition*, 14, 141-149. doi:10.3758/BF03198374
- Krueger, L. E., & Hallford, E. W. (1984). Why $2 + 2 = 5$ looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition*, 12, 171-180. doi:10.3758/BF03198431
- 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. doi:10.1037/0096-3445.125.3.284
- 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. doi:10.3758/BF03211420
- Maslany, A. J., & Campbell, J. I. D. (2013). Failures to replicate hyper-retrieval-induced forgetting in cognitive arithmetic memory. *Canadian Journal of Experimental Psychology*, 67(1), 72-77. doi:10.1037/a0031138
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679-690. doi:10.3758/s13428-010-0049-5
- Meagher, P. D., & Campbell, J. I. D. (1995). Effects of prime type and delay on multiplication priming: Evidence for a dual-process model. *Quarterly Journal of Experimental Psychology*, 48A, 801-821. doi:10.1080/14640749508401418
- Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 539-543. doi:10.1037/0096-1523.14.3.539
- Nathoo, F. S., & Masson, M. E. J. (2016). Bayesian alternatives to null-hypothesis significance testing for repeated-measures designs. *Journal of Mathematical Psychology*, 72, 144-157. doi:10.1016/j.jmp.2015.03.003

- Núñez-Peña, M. I., Gracia-Bafalluy, M., & Tubau, E. (2011). Individual differences in arithmetic skill reflected in event-related brain potentials. *International Journal of Psychophysiology*, *80*(2), 143-149. doi:10.1016/j.ijpsycho.2011.02.017
- Romero, S. G., Rickard, T. C., & Bourne, L. E. (2006). Verification of multiplication facts: An investigation using retrospective protocols. *The American Journal of Psychology*, *119*, 87-120. doi:10.2307/20445320
- Rotem, A., & Henik, A. (2013). The development of product parity sensitivity in children with mathematics learning disability and in typical achievers. *Research in Developmental Disabilities*, *34*(2), 831-839. doi:10.1016/j.ridd.2012.11.001
- Rotem, A., & Henik, A. (2015). Development of product relatedness and distance effects in typical achievers and in children with mathematics learning disabilities. *Journal of Learning Disabilities*, *48*(6), 577-592. doi:10.1177/0022219413520182
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356-374. doi:10.1016/j.jmp.2012.08.001
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237. doi:10.3758/PBR.16.2.225
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA, USA: Psychology Software Tools.
- Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. (1982). A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 320-335. doi:10.1037/0278-7393.8.4.320
- Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, *40*, 827-843. doi:10.3758/s13421-012-0211-7
- Szűcs, D., & Soltész, F. (2010). Event-related brain potentials to violations of arithmetic syntax represented by place value structure. *Biological Psychology*, *84*(2), 354-367. doi:10.1016/j.biopsycho.2010.04.002
- van der Ven, S. H. G., Straatemeier, M., Jansen, B. R. J., Klinkenberg, S., & van der Maas, H. L. J. (2015). Learning multiplication: An integrated analysis of the multiplication ability of primary school children and the difficulty of single digit and multidigit multiplication problems. *Learning and Individual Differences*, *43*, 48-62. doi:10.1016/j.lindif.2015.08.013
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi:10.3758/BF03194105
- Wetzels, R., van Ravenzwaaij, D., & Wagenmakers, E.-J. (2015). Bayesian analysis. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1-11). doi:10.1002/9781118625392.wbecp453
- Zbrodoff, N. J., & Logan, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 83-97. doi:10.1037/0278-7393.16.1.83
- 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, USA: Psychology Press.