Journal of Applied Research in Memory and Cognition 2 (2013) 25–32

Contents lists available at SciVerse ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

A test of two methods of arithmetic fluency training and implications for educational practice

Drew Walker[a], Laura Mickes[a], Daniel Bajic[a], Charles R. Nailon[b], Timothy C. Rickard[a,*]

[a] University of California, San Diego, United States
[b] Spring Valley Community Center, Spring Valley, United States

ARTICLE INFO

Article history:
Received 1 October 2012
Received in revised form 25 January 2013
Accepted 3 February 2013
Available online 11 February 2013

Keywords:
Arithmetic
Skill
Education
Mathematical cognition
Testing

ABSTRACT

Children are exposed to multiple training tasks that are intended to support acquisition of basic arithmetic skills. Surprisingly, there is a scarcity of experimental research that directly compares the efficacy of those tasks, raising the possibility that children may be spending critical instructional time on tasks that are not effective. We conducted an experiment with 1st through 6th grade children comparing two arithmetic training tasks that are widely used: answer production training and fact triangle training. Results show that answer production training produces substantial fluency gains, whereas fact triangle training does not. Further, we show that, despite theoretical considerations that suggest otherwise, fact triangle training does not produce more flexibly applicable learning. Implications for memory representation, arithmetic fluency training, and broader educational strategy are discussed.

Elementary school children are often exposed to multiple training tasks for learning a target skill. In the domain of single-digit arithmetic, the focus of this paper, answer production practice (i.e., drill) is often supplemented by completing ordered multiplication tables, as well as by more recently introduced tasks such as arithmetic fact triangle exercises (to be described below) and various hybrid exercises that involve both answer production and study-like activities that are intended to be entertaining (e.g., the "color-by-numbers" task in which the answer to an arithmetic problem indicates the color to be used for filling in a given section of a hidden picture). Similar strategies are used in other domains. In the case of spelling, for example, having students write or recite the spelling of aurally presented words (i.e., testing with feedback) is often supplemented by tasks in which the words (and hence the spellings) remain visually available while students perform some study activity, such as rainbow writing (i.e., writing words in multiple colors), alphabetizing, typing the words, and building word search puzzles.

Remarkably, it appears that none of the more recently introduced training tasks have been tested for efficacy, either against each other or against more traditional approaches. As argued by Daniel (2013), introduction of new tasks prior to empirical vetting in an authentic educational setting raises the possibility that children spend substantial time performing tasks that are inefficient in producing the target learning. In the extreme, it is possible that some tasks produce little or no useful learning.

In the case of arithmetic, a number of studies have documented the effectiveness of answer production (AP) training, and progress has been made toward identifying optimal instantiations of that approach (Burns, 2004, 2005; Cooke & Reichard, 1996; Cooke, Guzaukus, Pressley, & Kerr, 1993). We are aware of no prior research, however, in which efficacy was evaluated either among the other classes of training tasks or between those tasks and AP training. Below we report, for the first time in the literature, results of an experiment that compares the relative effectiveness of two commonly used tasks: (1) a form AP training involving production of answers to a series of randomly sequenced problems, and (2) the more recently introduced arithmetic fact triangle exercises.

Fact triangles (see Fig. 1) are created from triplets of numbers (e.g., 4, 7, 28) that are related through complementary operations (i.e., addition/subtraction or multiplication/division). Fact triangle practice takes various forms, but often involves visual presentation of the triangle, with students vocally rehearsing, or writing, the corresponding family of arithmetic problems (e.g., $4 \times 7 = 28$; $7 \times 4 = 28$; $28/7 = 4$; $28/4 = 7$). Unlike AP training, there is no production of an answer from memory. Rather, all of the numbers are presented in the triangle. The student's task is to recall which sequences of numbers constitute arithmetic problems that correspond to the triangle and to write those numbers in the correct sequence in the underlined spaces below the triangle.

Fact triangle exercises have been used broadly in multiple curricula over at least the last decade. Variants of these exercises are included as part of several widely adopted elementary mathematics

* Corresponding author. Tel.: +1 858 822 0122.
E-mail address: trickard@ucsd.edu (T.C. Rickard).

2211-3681/$ – see front matter © 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jarmac.2013.02.001

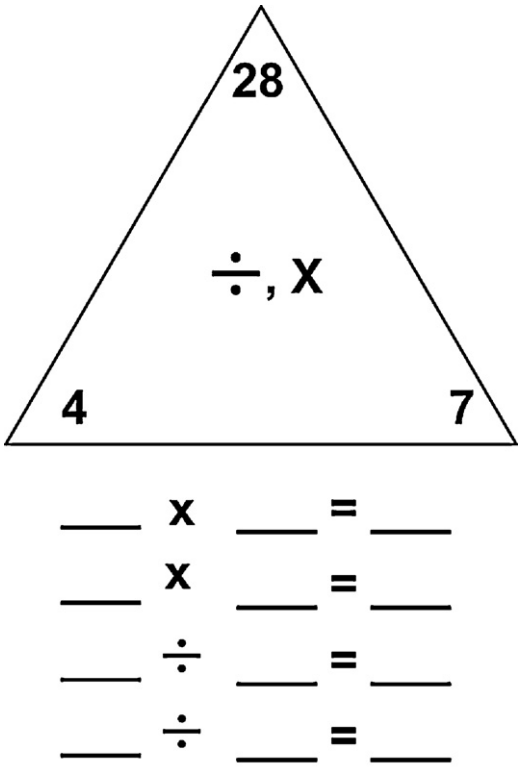

Fig. 1. An example fact triangle for the number triplet 4, 7, 28.

curricula (e.g., Everyday Mathematics, Houghton-Mifflin Mathematics, McGraw-Hill Mathematics), and a web search confirms the availability of commercial materials targeted for home use. Given the likelihood that millions of children have or will perform these exercises as part of their arithmetic training, it is important to test their efficacy.

1. AP training vs. fact triangle exercises: theoretical considerations

AP training is known to produce a shift from use of an algorithm—defined here as the application of a series of calculation steps that, if applied correctly, are guaranteed to produce the correct answer—to direct retrieval of answers from memory (e.g., Logan, 1988; Rickard, 1997; Siegler, 1988). To solve a problem like 4×7, children may initially use a repeated addition algorithm ($7+7+7+7=28$), but, with sufficient practice, will shift to single-step retrieval of the answer from memory. Experiments with adults have shown that the shift to retrieval in similar tasks can yield abrupt, multi-second improvements in response time (e.g., Rickard, 2004). That shift is thus an important event in the development of fluent arithmetic performance.

Prior research has demonstrated that the shift to retrieval through AP training is highly specific to practiced problems. Rickard, Healy, and Bourne (1994) formalized this specificity of learning in their identical elements (IE) model of arithmetic fact representation. According to that model, AP training results in a separate memory representation for each unique combination of stimulus and response elements, ignoring superficial factors such as operand order and stimulus modality. Thus, the model proposes three distinct problem answer associations in memory for each arithmetic number triplet. For the example triplet 4, 7, 28, the associations can be written as: $(4, 7, x) \rightarrow 28$, $(28/7) \rightarrow 4$, and $(28/4) \rightarrow 7$. The IE model predicts that memory-based learning during AP training will transfer positively to untrained problems only if the elements of those problems and their stimulus-response roles fully match (ignoring the superficial factors) the stimulus and response elements of one of the trained problems. Hence, practice-based improvements in performance for the problem 4×7 would transfer to 7×4, but would not transfer to the complement division problems (e.g., 28/7). Nor would learning the division form of a problem transfer to multiplication. Further, learning for one division problem (28/4) does not transfer to its complement (28/7), because the stimulus and response elements of those problems do not fully match. For both adults and children, the results of multiple experiments are broadly consistent with the IE model for the cases of both multiplication–division and addition–subtraction (Bajic, Kwak, & Rickard, 2011; Campbell & Agnew, 2009; Rickard, 2005; Walker, Bajic, Kwak, Mickes, & Rickard, under review).

AP training closely approximates the educational target skill of producing answers to randomly encountered problems. Fact triangle exercises, in contrast, do not approximate a target skill; it is safe to assume that most students will not encounter fact triangles beyond elementary school. Thus, from a transfer appropriate processing perspective (Morris, Bransford, & Franks, 1977), AP training should be superior for acquiring the target skill. Nevertheless, two considerations raise the distinct possibility that fact triangle training may: (1) promote equivalent or faster onset of fluent arithmetic answer production relative to the same amount of AP training, (2) promote a deeper understanding of the material, and (3) produce more flexibly applicable memory representations.

First, throughout each fact triangle trial, the full number triplet is perceptually available, and thus memory encoding of that triplet can, in principle, occur continuously. In contrast, learning through AP practice involves a transition from slow algorithms to fast memory retrieval. When the algorithm is used, there can be no learning of the problem–answer association until the answer has been generated by the algorithm, and the algorithm constitutes the majority of the time that children spend attending to the problem. Indeed, it is not clear that children spend any time attending to an arithmetic problem in an AP training context after generating the answer. Rather, they may shift their attention almost immediately to the next problem. As such, the percentage of time-on-task in which arithmetic fact memorization occurs on algorithm AP trials is likely low.

A second factor that appears to favor fact triangle exercises is that they may promote development of a single, holistic memory representation for each arithmetic number triplet. Once developed, that representation may be accessible for finding the answer to any arithmetic problem that involves the same number triplet (for related theoretical perspectives, see Anderson, Fincham, & Douglass, 1997; Campbell & Agnew, 2009; Rabinowitz & Goldberg, 1995). For example, a holistic representation of the triplet 4, 7, 28 should be able to support performance on $4 \times 7 =$ __, $7 \times 4 =$ __, $28/7 =$ _, and $28/4 =$ _. In contrast, the shift to retrieval through AP training appears to require learning of three distinct IE associations to support memory-based answer production for all problems corresponding to each number triplet (see prior discussion). These two advantages combined—higher percentage of time-on-task in which memory encoding can occur, and the possible formation of holistic rather than IE memory representations—may afford a significant advantage for fact triangle training, at least during the intermediate stages of arithmetic fluency development wherein the shift to memory-based performance is occurring.

2. Methods

2.1. Setting and subjects

The experiment was conducted as part of an after-school math club program held at the Spring Valley Community Center in Spring

Valley, California. The Math Club recruits students from all schools within the Spring Valley district. Within that district, 75% of students are enrolled in the free or reduced lunch program and only 57% are categorized at the basic-or-above level of mathematics. The ethnicity distribution of that school district is 54.7% Hispanic, 22.5% White not Hispanic, 11.3% African American, 6.4% Asian, 2.4% two or more races, 1.2% Filipino, and .7% American Indian.

The math club met from 3:30 pm to 5:00 pm on Wednesdays and Thursdays for each of six weeks during the Fall of 2009. The experiment took place over the course of seven 40-min sessions, beginning on Thursday of the second week of Math Club and ending on Thursday of the fifth week. The study was conducted in a multipurpose room with children sitting at tables organized by grade and working independently using paper and pencil. All students provided voluntary, written assent to participate and all parents provided voluntary written consent.

Sixty-one second through sixth grade students participated in at least one experimental session. Seventeen of those students stopped attending Math Club (i.e., not just our study but Math Club generally) at the onset of a two week district school break that occurred during the third and fourth weeks, and an additional 9 students did not attend the posttest session, yielding 35 students who attended the posttest session and at least one training session.[1] Nineteen of those students also attended to pretest session.

2.2. Materials, design and procedures

Each student was trained and tested on one of four problem sets (see Appendix A): "easy" addition–subtraction (set 1; second graders), "difficult" addition–subtraction (set 2; third graders), "easy" multiplication–division (set 3; fourth graders), or "difficult" multiplication–division (set 4; fifth and sixth graders). The goal was to assign each student, based on grade level, a problem set on which he or she was expected to have basic procedural competence (i.e., problems that they were capable of solving using the calculation algorithm without systematic error, but for which performance was unlikely to reflect highly automatic memory-based performance). This design assured that students could work independently with minimal risk of acquiring error associations through repeated incorrect calculation, while also setting the stage for observing substantial learning through training.

The number triplets corresponding to each set were divided into two subsets (subsets A and B, as shown in Appendix A). These subsets were balanced on solution size (i.e., sum or product). Within each grade-level, half of the students received subset A as their AP trained problems, and subset B as their fact triangle trained problems. The other half of students received the opposite subset assignment. Due to the subject attrition noted earlier, the counterbalancing described here and elsewhere in Section 2 was not fully realized. However, the pretest results for number correct (described later) were nearly identical for AP and fact triangle conditions. It is therefore unlikely that incomplete counterbalancing substantially biased either the posttest or the pre–post results.

The experimental design involved three phases: a pretest (session 1), a multi-session training phase (sessions 1 through 6), and a posttest (identical to the pretest) and transfer test phase (session 7). The pretest, posttest and transfer test all involved answer production, the practically important target skill explored here. Training condition (fact triangle training and AP training) was manipulated within subjects.

During all phases of the experiment students worked blocks of problems contained in a binder using paper and pencil. To begin each timed block of problems, the experimenter held up a stopwatch and said "Ready, set go!" at which time the students turned a cover page and began working. At the end of each block, the instructor said "Stop now, please," at which point students were instructed to put their pencils down immediately. There was a 2-min break between blocks within a session, during which students were told to thumb forward to the cover page that identified the beginning of the next part (block), but not to proceed until instructed. Assistants confirmed student compliance.

During all training and testing blocks, students were instructed to work as quickly and accurately as possible, writing answers in the underlined spaces for each problem. More problem sheets were included for each block than students were able to complete. This aspect of the design served two purposes. First, it allowed the number of problems completed during each timed block to be used as a measure of performance rate. Second, because all students worked throughout each block (i.e., no students finished early), students who worked slowly were not exposed to direct performance evaluation from students who worked quickly.

2.2.1. Training

During the first session, students were given the pretest (described below) followed by four training blocks. In each of the five remaining training sessions, students received six training blocks. Half of the students received AP training on odd-numbered blocks within each session and fact triangle training on even-numbered blocks, with that order reversed for the remaining students. Each training block lasted for 3.5 min.

During AP training, students worked sequentially through pages of problems, with 12 problems per page, presented in horizontal format. Only one problem from each of the 12 number triplets in the student's problem set was presented per page, randomly ordered. On each page, half of the problems were presented in one operation (e.g., addition or multiplication) and the other half were presented in the complement operation (subtraction or division), determined randomly. Each student received a different random sequence of problems across each consecutive set of four pages, subject to the constraints above. Problems were printed only on the front side of each page, so that students were unable to view their prior work as a short cut to answer production.

During fact triangle training, students also worked sequentially through pages of problems, with 12 fact triangle problems per page (see Fig. 1 for an example problem), one corresponding to each of the 12 number triplets. These sheets of the same 12 problems repeated throughout each block, with fact triangles randomly ordered anew on each sheet. For each fact triangle, students wrote the appropriate numbers in each of the underlined spaces below the triangle (e.g., _ × _ = __).

All students used the Everyday Mathematics curriculum in school, which familiarizes them with fact triangle exercises starting from the first grade. We also gave all students a brief tutorial on fact triangles in session 1, prior to the experiment. The students did not exhibit any conceptual or systematic procedural difficulties with the fact triangle problems, nor with the AP training or testing.

[1] Note that the student attrition is unlikely to be problematic with respect to our goal of differentiating the effectiveness of the two training tasks, for two reasons. First, the cessation of attendance during the district break was presumably the result of family vacations or other activities. Given that all students received both fact triangle training and AP training (i.e., given that training task was manipulated within-subjects), it cannot be the case that those students stopped attending based on which task was encountered. Second, although 9 students did not attend the posttest session, there was no disproportionate lack of attendance for that session. Rather, on any given day of training or testing, some students did not attend Math Club, presumably due to influences that, from the experimenter's perspective, can be considered random (e.g., illness or competing family activities). In light of the fact that students were not aware that the posttest session would involve only AP exercises, it is also not possible that some students did not attend the posttest selectively because of their differential performance on or liking of the AP vs. fact triangle tasks.

Table 1
Number of students trained on each stimulus set per grade.

Grade	Stimulus set	Students
2	Addition–subtraction 1	4
3	Addition–subtraction 2	11
4	Addition–subtraction 2	5
4	Multiplication–division 1	3
5	Multiplication–division 2	6
6	Multiplication–division 2	2
Total		31

2.2.2. Testing and transfer

The pre- and posttests were identical and involved the same procedures as the AP training blocks. Both tests involved two answer production blocks, one block for problems trained in the AP condition, and another for problems trained in the fact triangle condition (with block-order counterbalanced over students). Each pretest block lasted for 5 min. Each posttest block lasted 3.5 min.[2] The transfer test also involved two 3.5 min AP blocks, one block for problems from each training condition. On the transfer test blocks, standard format AP problems were mixed with atypical format problems (i.e., for the triplet 4, 7, 28, the following problems were on the transfer test 4 × 7 = _; 4 × _ = 28; _ × 4 = 28; _ = 4 × 7; 28 = _ × 7; 28 = 4 × _; 28/7 = _; 28/4 = _; 28/_ = 7; _ = 28/7; 4 = _/7; 4 = 28/_). Each page of the transfer test included one problem from each number triplet and one problem from each of the 12 formats, randomly determined for each student. Over 12 consecutive pages of transfer problems, each number triplet would be presented in each of the 12 formats, although no students approached completion of that full set in the time allowed. The transfer test was designed to provide insight into the relative flexibility of learning as acquired through AP vs. fact triangle training.

3. Results

As noted earlier, 35 students attended at least one training session and the posttest/transfer test session. Two of those students were excluded as outliers due to very high error and skipped problem rates and two were excluded as outliers because they correctly answered fewer than 12 problems on the two posttests block combined. (Inclusion of these four subjects in the analyses did not alter the patterns of significance described below.) The 31 remaining students correctly answered at least 24 posttest problems. Number of students in each grade level, and trained under each stimulus set, is shown in Table 1. Because Math Club enrollment was not under our control, sample size varied by grade, and hence also by stimulus set assignment.

The number of problems skipped on each pre, post, and transfer test block was very low, averaging .58 problems in the fact triangle condition and .52 problems in the AP condition. These skipped problems were excluded from the data analyses reported below. Results were nearly identical in supplementary analyses in which skipped problems were counted as errors.

3.1. Training

The mean number of training sessions attended for the 31 students was 4.6. Because the same number of fact triangle and AP blocks were given in each session, all students received identical training time on the two tasks. Preliminary analyses indicated that posttest performance did not depend significantly on number of training sessions attended, and hence that factor will not be considered further. The mean number of correctly solved fact triangle problems over training session 2 through 6 was 35.1, 36.75, 41.3, 40.5, and 37.3 problems, respectively (AP training effects were not analyzed, but are evident in the pre–post results described below). Those results confirm the observation of research staff that students attended to the fact triangle problems throughout training, with no noticeable fall-off in intensity relative to AP training blocks.

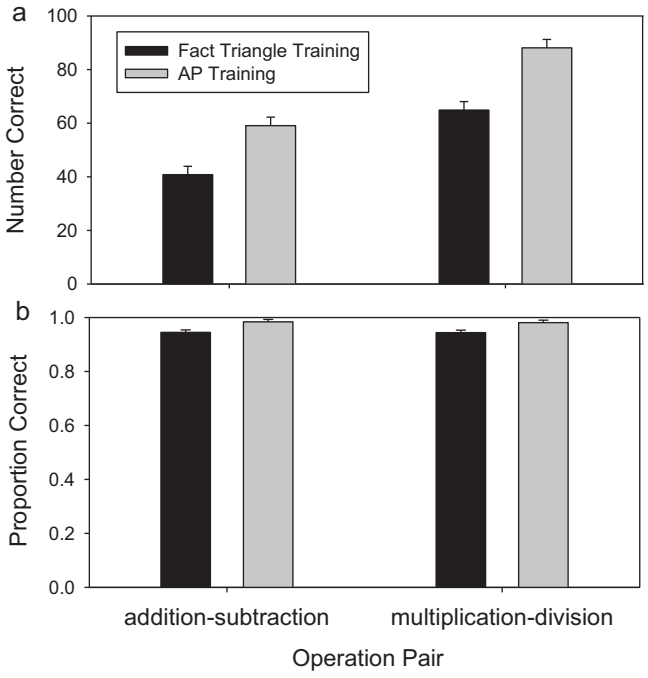

Fig. 2. Posttest results by operation-pair and training task. (Panel a) Number of problems correctly solved out of total number attempted; (Panel b) proportion correct out of total number attempted. Error bars represent the standard error based on the within-subjects error term of the ANOVA (see Loftus & Masson, 1994).

3.2. Posttest results

27 of the 31 students solved more AP trained than fact triangle trained problems on the posttest. Results are shown in Fig. 2a as a function of operation-pair (addition–subtraction, $n = 20$, and multiplication–division, $n = 11$) and training task. A factorial Analysis of Variance (ANOVA) on those factors (with operation-pair as a between subjects factor and task a within subjects) confirmed a strong advantage for AP training, $F(1, 29) = 22.9$, $p < .0001$, η_p (partial eta-squared) = .44, and a trend toward better performance in the multiplication–division condition, $F(1, 29) = 3.62$, $p = .07$, $\eta_p = .11$, but no significant interaction, $F(1, 29) = .31$, $p = .31$, $\eta_p = .01$.[3] For proportion correct (Fig. 2b), there was again a significant advantage for AP training, $F(1, 27) = 10.1$, $p = .004$, $\eta_p = .276$, but there were no effects for either operation-pair or the interaction (both F's < 1.0; η_p's < .01). With respect to both number correct and accuracy, the numerical performance advantage for AP training held at each grade level.

3.3. Pre–post results

The posttest results demonstrate that AP training produced more learning, but they do not address the question of how much

[2] Due to experimenter error, the duration of all posttest and transfer blocks remained at 3.5 min (as it was by design for the training blocks) instead of the intended 5-min duration that was used for the pretest.

[3] Analyses were conducted using SAS Proc GLM and Type III sums of squares, which adjusts for unequal sample sizes.

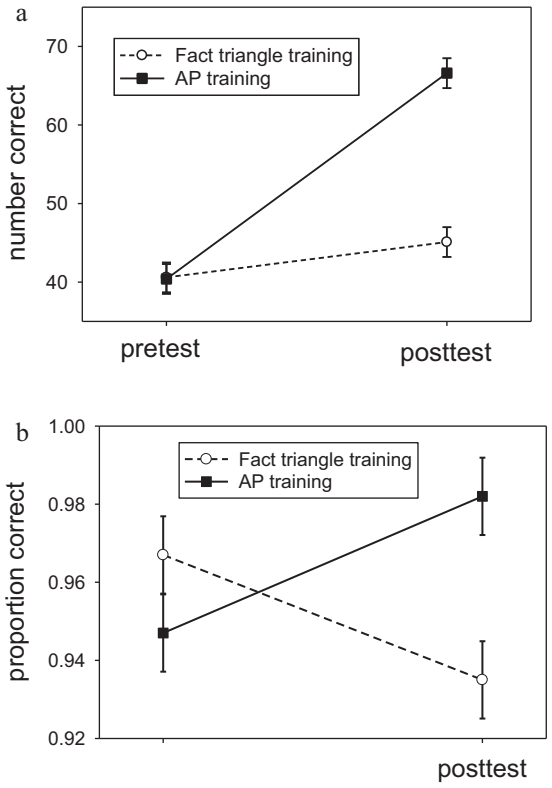

Fig. 3. Pre- and posttest results. (Panel a) Number of problems correctly solved out of total number attempted; (Panel b) proportion correct out of total number attempted. Error bars represent the standard error based on the within-subjects error term of the ANOVA (see Loftus & Masson, 1994).

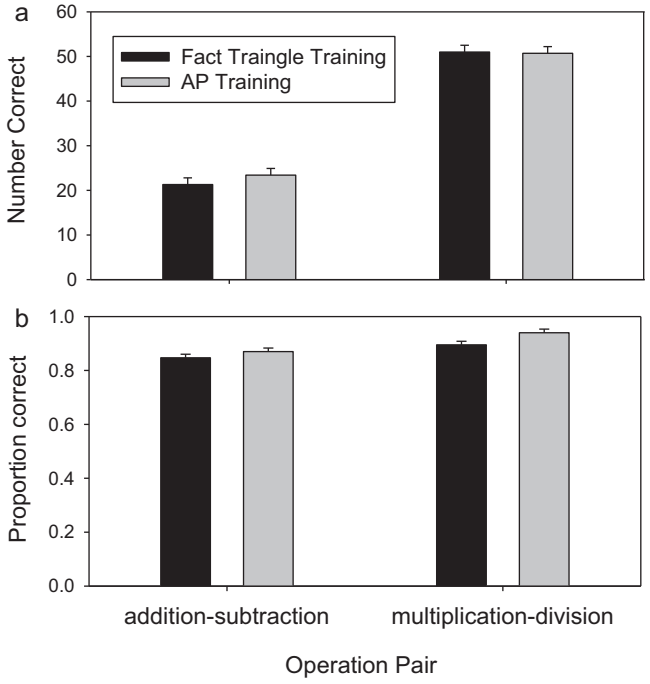

Fig. 4. Transfer test results by operation-pair and training task. (Panel a) Number of problems correctly solved out of total number attempted; (Panel b) proportion correct out of total number attempted. Error bars represent the standard error based on the within-subjects error term of the ANOVA (see Loftus & Masson, 1994).

learning occurred in the two training conditions. The pre–post results for the 19 students who attended the pretest address that question. Means, averaged over the 19 students, are shown in Fig. 3 for number correct (panel a) and proportion correct (panel b).[4] For number correct, there was 65% pre–post improvement in the AP training condition but only an 11% improvement in the fact triangle training condition.

A 2 (training task) by 2 (test: pre vs. post) within subjects ANOVA was performed on number correct (collapsed over the factor of operation-pair). The results confirmed the effects of training task, $F(1, 18) = 11.4$, $p = .003$, $\eta_p = .48$, test, $F(1, 18) = 16.5$, $p < .001$, $\eta_p = .39$, and their interaction, $F(1, 18) = 32.4$, $p < .0001$, $\eta_p = .64$.

The simple pre–post learning effect was highly significant for the AP condition, $t(1, 18) = 4.89$, $p < .001$, d (Cohen's d) = .86, and exhibited a non-significant trend in the fact triangle condition, $t(1, 18) = 1.69$, $p = .10$, $d = .47$. These analyses confirm the superior learning in the AP condition, but also raise the possibility that fact triangle training produced modest but undetected learning that was transferable to the AP task. Alternatively, the pre–post effect for fact triangles may reflect school learning that occurred during the five weeks of the experiment.

For proportion correct, there was again pre–post improvement in the AP condition, but there was performance worsening in the fact triangle condition, an interaction that was confirmed by the AVOVA, $F(1, 18) = 11.1$, $p < .004$, $\eta_p = .38$. Neither the main effect of Test, $F(1, 18) = .03$, $\eta_p = <.01$, nor training task, $F(1, 18) = 1.5$, $p = 24$, $\eta_p = .077$, were significant.

3.4. Transfer results

Transfer results are shown by grade and training task in Fig. 4. The AVOVA on number correct (panel a) indicated a significant effect of operation-pair, $F(1, 29) = 7.94$, $p = .009$, $\eta_p = .22$, but no effects of either training task or the interaction (both F's < 1.0; η_p's > .02). For proportion correct (panel b), there were marginally significant trends toward an advantage for multiplication–division, $F(1, 29) = 3.02$, $p = .056$, $\eta_p = .12$, and the AP training task, $F(1, 29) = 3.97$, $p = .093$, $\eta_p = .89$, but no trend toward an interaction, $F(1, 29) < 1.0$, $\eta_p < .01$. In supplemental analyses on proportional correct in which trained problems were eliminated (leaving only the atypical format problems), the trend toward an AP advantage was eliminated, $F(1, 29) = 1.26$, $p = .27$, $\eta_p < .01$. Thus there is no evidence that AP training produced better transfer to atypical format problems than did fact triangle training.

4. Discussion

Across multiple grades and for the operation pairs of both addition–subtraction and multiplication–division, fact triangle training yielded at best minimal improvements in answer production, whereas AP training yielded a highly significant 65% improvement. On the transfer test to atypical format problems, there were no significant differences between the two training conditions.

These results suggest that fact triangle training does not generally result in the formation of holistic memory representations for arithmetic number triplets, at least not in a form that is accessible for later answer production. This finding may seem counterintuitive in light of the fact that episodic memory can likely take the

[4] For the pre–post comparison in Fig. 3a, the number of correctly solved problems on the 5 min pretest was adjusted by multiplying each subject's result by 3.5/5, yielding the expected value had the pretest duration been the same as the posttest duration (3.5 min). Assuming a roughly constant rate of problem solving over the 5-min pretest, this adjustment does not introduce a bias in the pre–post main effect, and it is in our judgment very unlikely to have materially affected the critical interaction between test and training condition. No timing adjustment was made to the posttest or transfer test results in any of the figures or analyses.

form of holistic representation. It seems intuitive, for example, that extensive study of a small set of number triplets would create long-term memories that are holistic and that could support retrieval-based answer production for all corresponding arithmetic facts. Our results, however, suggest that if such memories are formed during fact triangle study, they are unstable and quickly forgotten. It may be that IE representations, which may form as a direct consequence of answer production rather than through study or other generative activities, are more akin to a procedural than episodic-holistic representations and are more resistant to forgetting (for further discussion see Bajic et al., 2011; Rickard & Bajic, 2006). If so, then these results are in accord with the procedural reinstatement principle of Healy and colleagues (e.g., Lohse & Healy, 2012), according to which procedural knowledge is well retained but highly specific (recall the specificity claims of the IE model), whereas as declarative knowledge (i.e., memorized number triplets) is more flexible but poorly retained.

Although the observed specificity of learning (i.e., poor performance on atypical problems even in the AP condition) is disappointing from an applied perspective, it may underestimate the AP transfer that occurs once children have learned the solve-for-*x* arithmetic procedure (e.g., $4 + X = 7 \rightarrow X = 7 - 4$). When faced with an atypical format problem (e.g., $4 + \_\_ = 7$), children with that skill may be able to treat the blank space as, effectively, a variable and use the solve-for-*x* procedure to transform the problem into a familiar format. At that point, prior AP training (e.g., on the problems $7 - 4$) would be expected to facilitate answer production. Hence, it is plausible that AP training does yield better transfer than does fact triangle training at a later point in mathematical skill development.

4.1. Relations to testing, transfer appropriate processing, and the generation effect

It may be tempting to frame the advantage for AP training in terms of the well-established testing effect (i.e., the broad finding in primarily verbal domains that testing is superior to study as a training method for a later test; e.g., Roediger & Karpicke, 2006a, 2006b). By that account, AP training is a form of testing whereas fact triangle training is construed as a form of study. Fact triangles, however, are perhaps better characterized as an alternative form of testing; subjects need not produce a numerical answer, but they must produce from memory the correct ordering of the numbers in the spaces provided below the triangle. Hence, the analogy to study vs. testing comparisons in the broader literature is tenuous. A better match to the testing effect literature would involve comparing AP training (e.g., $4 \times 7 = \_\_$) to pure study of presented arithmetic problem–answer combinations (e.g., $4 \times 7 = 28$).

Our results are consistent, however, with the more specific hypothesis that the testing effect reflects retrieval practice of the critical response (e.g., Karpicke & Roediger, 2008). The results further speak to the robustness of the retrieval practice effect, extending it to arithmetic and to the case of extensive training among children (for a demonstration of retrieve practice among children without extended training, see Reodiger, Agarwal, McDaniel, & McDermott, 2011). The results also show that retrieval practice can be more effective than alternative forms of instruction even when the percent time-on-task available for memory encoding is relatively small, as is the case for AP algorithm trials relative to fact triangle trials.

Our results are also consistent with the transfer appropriate processing (TAP) principle (Morris et al., 1977), according to which training will be most effective when the processing during training matches that during test. From that perspective, it is possible that testing on fact triangles rather than on answer production would have yielded the opposite outcome: better test performance following fact triangle than following AP training. As noted earlier, however, fact triangle proficiency is not a target education skill, and hence the applied implications of such a finding would be negligible. Also, given that both tasks are currently used in school curricula, the consistency of our results with TAP does not diminish their practical import. From a theoretical perspective, it was not obvious a priori whether the TAP principle would hold for these tasks in light of the two factors discussed earlier that appeared to substantially favor fact triangle training, Our study bears some similarities to the work on the generation effect (Slamecka & Graf, 1978), the finding that an active generation process (e.g., word stem completion) results in better subsequent memory for that word than does simply reading the word. Fact triangle processing can be construed as a type of generation activity, and yet it produced at best a small improvement in answer production. Fact triangles differ from the typical generation conditions in numerous respects, however, and generation effects are known to be sensitive to conditions (e.g., mixed vs. pure list designs), so the apparent lack of a generation effect in this case is not entirely surprising. Finally, note that our findings are consistent with recent work showing that implicit generation (word stem completion; analogous to our fact triangle task) is less effective than is explicit recall practice (analogous to our AP task) in facilitating final recall (Karpicke & Zaromb, 2010).

4.2. Practical implications for instruction in arithmetic and other domains

The current experiment was conducted in an authentic educational setting, with age and skill appropriate children. The results thus have straightforward practical implications. Firstly, they suggest that, although fact triangles may have a place as a method for teacher-led discussion about the relations between complement operations (i.e., they may be effective in illustrated arithmetic principles such as operation complementarity and commutativity), they are not an effective vehicle for fluency training or for establishing flexibly applicable arithmetic skill. Although it is conceivable that alternative ways of implementing fact triangle training will yield learning that is superior to pure AP training (e.g., mixed training on both tasks), for the time being the data motivate a simple change in arithmetic instruction that may substantially improve learning in some curricula: deemphasize fact triangle exercises in favor of more AP training. It remains to be determined whether other forms of arithmetic instruction, such as multiplication table completion, are as effective as is AP training.

More tentatively, our results may have broader educational implications. They constitute, in effect, an "existence proof" that all approaches to teaching a target skill cannot be assumed to have equivalent efficacy. Simply exposing children to the to-be-learned material does not necessarily produce the target learning. Whereas that statement is not controversial from the perspective of experimental research on skill and memory, current educational practice in domains such as arithmetic and spelling suggest it is not always heeded in curriculum design and task selection.

We suspect that one motivation for the introduction of multiple training tasks for a target skill is to promote student engagement with topics that might, using a single-task training approach, become tedious or boring (though we are aware of no evidence that speaks to that possibility). It is also possible that, to the extent that there are pronounced learning style or other individual differences among students (for review and critique, see Kozhevnikov, 2007; Pashler, McDaniel, Rohrer, & Bjork, 2008; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999) exposure to multiple tasks increases the likelihood that all students achieve at least some degree of learning. There is currently no compelling evidence for learning style effects on learning in any domain, however (see Pashler et al., 2008). With respect to the current experiment, 27 of the 31 students

who took the posttest exhibited better performance following AP training than following fact triangle training. The remaining four students (13% of the sample) might learn better with fact triangles, but the results for those students might also reflect random factors that masked a ubiquitous underlying learning advantage for AP training. For this particular pair of learning tasks, then, there is little reason for concern about negative learning style or other individual difference consequences of focusing student time on AP training. More generally, the current results illustrate the potential of within-student experimental comparisons to simultaneously identify the instructional method that is more effective overall and to provide insight into possible interactions with learning style or other individual difference factors.

Our results point to the need for a broader research effort to establish the efficacy of popular training tasks in both arithmetic and other domains, particularly at the elementary school level where the introduction of untested training tasks appears to occur most frequently. In the interim, two broad guidelines for selecting tasks that are likely to produce learning can be followed: (1) drawing on the general literature on retrieval practice effects noted earlier (e.g., Karpicke & Roediger, 2008), tasks that require answer production are likely to be more effective than those that involve some form of study of other generative activity, and (2) training tasks that closely approximate an important skill that will later be needed (in school or in everyday life) are preferable to those that do not. Fact triangle training is an example of a task that meets neither of those criteria. There are no doubt exceptions to these guidelines, but in our view empirical validation (preferably using randomized control designs) of tasks that do not meet both criteria is needed before they should be considered for introduction into the classroom or as homework.

Acknowledgments

We thank the Spring Valley Community Center for graciously allowing us to conduct this resource intensive multi-week research project. We thank the Math Club students and their parents for participating in this study. We thank Katy Brecht, Diana Graham, and Jennifer Guerrero for their assistance in materials development, study execution, and data tabulation.

Appendix A.

The number triplet from which problem sets were derived.

Set 1	Set 2	Set 3	Set 4
Subset A			
2 3 5	3 7 10	2 3 6	3 7 21
2 4 6	3 9 12	2 6 12	3 9 27
2 7 9	4 4 8	2 8 16	4 5 20
2 8 10	4 7 11	3 3 9	4 6 24
3 4 7	4 9 13	3 5 15	4 8 32
3 5 8	5 6 11	3 6 18	5 5 25
3 8 11	5 8 13	3 8 24	5 8 40
3 9 12	6 7 13	4 4 16	6 6 36
4 5 9	6 8 14	4 6 24	6 8 48
4 8 12	7 7 14	4 9 36	7 8 56
5 5 10	8 8 16	5 5 25	8 8 64
5 8 13	8 9 17	5 8 40	8 9 72
Subset B			
2 5 7	3 8 11	2 4 8	3 9 24
2 6 8	4 5 9	2 5 10	4 4 16
2 9 11	4 6 10	2 7 14	4 7 28
3 3 6	4 8 12	2 9 18	4 9 36
3 6 9	5 5 10	3 4 12	5 6 30
3 7 10	5 7 12	3 7 21	5 7 35
4 4 8	5 9 14	3 9 27	5 9 45
4 6 10	6 6 12	4 5 20	6 7 42
4 7 11	6 9 15	4 7 28	6 9 54
4 9 13	7 8 15	4 8 32	7 7 49
5 6 11	7 9 16	5 6 30	7 9 63
5 7 12	9 9 18	5 7 35	9 9 81

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 923–945.

Bajic, D., Kwak, J., & Rickard, T. C. (2011). Specificity of learning through memory retrieval practice: The case of addition and subtraction. *Psychonomic Bulletin & Review*, *18*, 1148–1155.

Burns, M. K. (2004). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education*, *25*, 167–175.

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, *28*, 237–249.

Campbell, J. I. D., & Agnew, H. (2009). Retrieval savings with nonidentical elements: The case of simple addition and subtraction. *Psychonomic Bulletin & Review*, *16*(5), 938–944.

Cooke, N. L., Guzaukas, R., Pressley, J. S., & Kerr, K. (1993). Effects of using a ratio of new items to review items during drill and practice: Three experiments. *Education and Treatment of Children*, *16*, 212–234.

Cooke, N. L., & Reichard, S. M. (1996). The effects of different interspersal drill ratios on acquisition and generalization of multiplication and division facts. *Education & Treatment of Children*, *19*, 124–142.

Daniel, D. (2013). Promising principles: Translating the science of learning to educational practice. *Journal of Applied Research in Memory and Cognition*, *1*, 251–253.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966.

Karpicke, D. J., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal or Memory and Language*, *62*, 227–239.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin*, *133*, 464–481.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

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

Lohse, K. R., & Healy, A. F. (2012). Exploring the contributions of declarative and procedural information to training: A test of the procedural reinstatement principle. *Journal of Applied Research in Memory and Cognition*, *1*, 65–72.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*, 105–119.

Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. Weinert, & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 225–242). Hillsdale, NJ: Erlbaum.

Reodiger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal or Experimental Psychology: Applied*, *17*, 382–395.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–311.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 65–82.

Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 250–257.

Rickard, T. C., & Bajic, D. (2006). Cued recall from image and sentence memory: A shift from episodic to identical elements representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 734–748.

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

Roediger, H. L., & Karpicke, J. P. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L., & Karpicke, J. D. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skills. *Journal of Experimental Psychology: General*, *117*, 258–275.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment*, *15*, 1–11.

Walker, D., Kwak, J., Bajic, D., Mickes, L., & Rickard, T.C. Specificity of children's arithmetic learning, under review.