Early Childhood Research Quarterly 30 (2015) 93–105

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

The impact of highly and minimally guided discovery instruction on promoting the learning of reasoning strategies for basic add-1 and doubles combinations

Arthur J. Baroody[a,*], David J. Purpura[b], Michael D. Eiland[c], Erin E. Reid[d]

[a] College of Education, University of Illinois at Urbana-Champaign, Champaign, IL 61820, United States
[b] Human Development and Family Studies, Purdue University, 1202 W. State Street, Room 231, West Lafayette, IN 47907-2055, United States
[c] Wisconsin Center for Education Research, University of Wisconsin-Madison, 1025 W. Johnson Street, Room 956, Madison, WI 53706-1796, United States
[d] Early Math Collaborative, Erikson Institute, 451 N LaSalle St, Chicago, IL 60654, United States

ARTICLE INFO

Article history:
Received 6 March 2013
Received in revised form 29 August 2014
Accepted 5 September 2014
Available online 16 September 2014

Keywords:
Discovery learning
Elementary mathematics
Computer-assisted instruction
Basic number combinations
Reasoning strategies

ABSTRACT

A 9-month training experiment was conducted to evaluate the efficacy of highly and minimally guided discovery interventions targeting the add-1 rule (the sum of a number and one is the next number on the mental number list) and doubles relations (e.g., an everyday example of the double 5 + 5 is five fingers on the left hand and five fingers on the right hand make 10 fingers in all) and to compare their impact with regular classroom instruction on adding 1 and the doubles. After pretest, 81 kindergarten to second-grade participants were randomly assigned to one of three training conditions: highly guided add-1 training, highly guided doubles training, or minimally guided add-1 and doubles practice. The highly guided add-1 training served as an active control for the highly guided doubles training and vice versa, and the minimally guided practice condition served to control for the impact of extra practice. ANCOVAs using pretest score and age as covariates indicated that both highly guided and minimally guided interventions were successful in promoting retention and transfer for the relatively salient add-1 rule, but only highly guided training produced transfer for the less-salient doubles relations. The findings indicate that the degree of guidance needed to achieve fluency with different addition reasoning strategies varies.

Introduction

The Common Core State Standards (CCSS; Council of Chief State School Officers [CCSSO], 2010) lay a framework for identifying the central skills and concepts pupils need to master at each grade level. CCSS Standard 6 in the grade 1 operations and algebraic thinking domain states: "Add and subtract within 20, demonstrating fluency for addition . . . within 10" and use reasoning strategies to determine sums. Fluency implies *efficient* (accurate and fast) production of sums. As used hereafter, the term also means *appropriate* and *adaptive* application of knowledge (e.g., selective application of a rule/strategy to novel problems not previously solved). Although there is general agreement that *all* children need to achieve fluency with basic sums (CCSSO, 2010; National Council of Teachers of Mathematics [NCTM], 2000, 2006; National Mathematics Advisory Panel [NMAP], 2008; National Research Council [NRC], 2001), there is disagreement about the best method(s) for achieving this goal. The main aim of this study was to gauge the efficacy of software designed to promote primary grade pupils' fluency with the most basic sums—the starting points of mental-addition fluency. A by-product of the research was comparing the relative efficacy of different instructional approaches as a step toward identifying best practices in mathematics education.

Instructional content of the interventions: why focus on reasoning strategies?

Reasoning strategies in general

The *meaningful learning* of a basic sum or family of basic sums entails three overlapping phases (Verschaffel, Greer, & De Corte, 2007). Initially, children typically use object or verbal counting to determine the sum (Phase 1: counting strategies). For example, for 2 + 3, a child might count, "Three, four is one more, *five* is two more—the answer is five." Then, as a result of discovering patterns or relations, children invent reasoning strategies, which they apply consciously and relatively slowly (Phase 2: deliberate reasoning

* Corresponding author at: Professor Emeritus of Curriculum & Instruction, College of Education, University of Illinois at Urbana-Champaign, Champaign, IL 61820, United States. Tel.: +1 217 586 5207; fax: +1 217 244 4572.
E-mail addresses: baroody@illinois.edu (A.J. Baroody), purpura@purdue.edu (D.J. Purpura), meiland@illinois.edu (M.D. Eiland), ereid@erikson.edu (E.E. Reid).

http://dx.doi.org/10.1016/j.ecresq.2014.09.003
0885-2006/

strategies). For example, children may discover that adding 1 is related to their existing knowledge of number-after relations. The discovery of this connection leads to the invention of a strategy, namely the *add-1 rule*. This reasoning strategy or rule specifies that the sum of any whole number, n, and 1 (or $1 + n$)—but not other items—is the number after n in the count sequence (e.g., the sum of 4 + 1 or 1 + 4, but not 4 + 0 or 3 + 4, is the number after *four—five*).

Learning reasoning strategies plays a critical role in the meaningful memorization of combinations (Phase 3: efficient, appropriate, and adaptive production of sums from a retrieval network) in two ways. One is that, with practice, reasoning strategies can become *automatic* (efficient and non-conscious; Jerman, 1970) and serve as a component of the retrieval system (Fayol & Thevenot, 2012). For instance, knowledge of the add-1 rule can be used to efficiently deduce any $n + 1$ or $1 + n$ combination, even previously unpracticed or multi-digit items, for which the child knows a number-after relation. The other way learning reasoning strategies can aid in achieving Phase 3 is that they provide children with an organizing framework for learning and storing both practiced and unpracticed combinations (Canobi, Reeve, & Pattison, 1998; Dowker, 2009; Rathmell, 1978; Sarama & Clements, 2009).

Most basic reasoning strategies

Research indicates that among the easiest sums for children to learn are the add-1 and doubles families (see reviews by Brownell, 1941; Cowan, 2003). Given the informal knowledge children bring to school, the add-1 family is a developmentally appropriate (as well as logical) place to begin mental-addition training. At the start of school, most pupils are so familiar with the count sequence they can fluently specify the number after a given number (Fuson, 1988, 1992). Achieving fluency with add-1 combinations simply entails connecting adding 1 to their extant number-after knowledge—that is, recognizing the add-1 rule (Baroody, 1989, 1992; Baroody, Eiland, Purpura, & Reid, 2012; Baroody, Eiland, Purpura, & Reid, 2013).

The doubles are also relatively easy to learn because they embody familiar real-world pairs of a set, such as a dog's two front legs and two back legs make four legs altogether (Baroody & Coslick, 1998; Rathmell, 1978). Using a familiar everyday situation to determine the sum of a double involves analogical reasoning, the simplest and most common method of reasoning. For example, if a carton of a dozen eggs has 12 eggs and each of the two rows of six eggs is analogous to 6 + 6, then the sum of 6 + 6 is 12 also. Another reason learning the doubles is relatively easy is that it can build on several common aspects of primary-level mathematics instruction (Baroody & Coslick, 1998). One is that the sums of doubles are all even numbers and parallel the even number (skip-count-by-two) sequence: "2, 4, 6, . . ." Another aspect is that the sum of a double is akin to the first two counts in various skip counts (e.g., 5 + 5 = 10 can be reinforced by knowing the skip-count-by fives: "five, ten").

The add-1 and doubles combination families are the basis for more advanced mental-addition reasoning strategies. For example, efficiently implementing the make-10 (e.g., 9 + 5 = 9 + 1 + 4 = 10 + 4 = 14) and near-doubles strategies (e.g., 5 + 6 = 5 + 5 + 1 = 10 + 1 = 11) requires fluency with the add-1 rule. Note that the near-doubles strategy also requires fluency with the doubles, such as 5 + 5 = 10 (Baroody et al., 2012, 2013).

Instructional method of the interventions: why guided discovery learning?

In an extensive review of the literature, the NRC (2001) concluded that Phase 2 can be accelerated by directly teaching reasoning strategies, *if* done conceptually. Direct teaching of reasoning strategies accompanied by explanation of their rationale is often recommended by mathematics educators (Rathmell, 1978; Thornton, 1978, 1990; Thornton & Smith, 1988) and utilized in many elementary curricula, such as *Everyday Mathematics* (University of Chicago School Mathematics Project [UCSMP], 2005).

However, not all conceptually based instruction is equally effective (Baroody, 2003). Chi (2009) hypothesized that constructive activities (producing responses that entail ideas that go beyond provided information) are more effective than active activities (doing something physically), which in turn are more effective than passive activities (e.g., listening or watching without using, exploring, or reflecting on the presented material). Direct instruction—even when it attempts to illuminate the rationale for a reasoning strategy—typically embodies passive activities. As a result, it may not actively engage many children, be comprehensible, or produce the adaptive expertise (meaningful learning) necessary to apply a strategy flexibly and appropriately (Hatano, 2003). For example, Murata (2004) found that Japanese children taught a decomposition strategy with larger-addend-first combinations did not exhibit strategy transfer when smaller-addend-first items were introduced. Torbeyns, Verschaffel, and Ghesquiere (2005) found that children taught the near-doubles strategy sometimes used the strategy accurately but other times inaccurately (e.g., relating 7 + 8 to 7 + 7 − 1 or 8 + 8 + 1 instead of 7 + 7 + 1 or 8 + 8 − 1).

Discovery learning may be better suited to learning basic reasoning strategies than direct instruction because it can involve active learning and constructive activities (Swenson, 1949; Thiele, 1938; Wilburn, 1949). Alfieri, Brooks, Aldrich, and Tenenbaum (2011) defined discovery learning as not providing learners with the targeted procedural or conceptual information but creating the opportunity to "find it independently . . . with only the provided materials" (p. 2). Discovery learning encompasses a wide range of methods, which may not be equally effective in all cases. At one extreme is *highly guided discovery*—well-structured and moderately explicit instruction and practice. Although a pattern, relation, or strategy is not explicitly provided or explained to a child (as in direct instruction), this type of discovery learning involves considerable scaffolding. Instruction and practice are organized to direct a child's attention to regularities or a strategy. For example, items are arranged sequentially to underscore a pattern or relation and prompts direct attention toward a regularity or strategy without explicitly stating it. Feedback provides some explanation of *why* a response is correct or incorrect as well as specifying whether an answer is correct or not. At the other extreme is *unguided discovery* (e.g., "free play"). With this type of discovery learning, children chose their own task, engage in unstructured activities, and do not receive adult feedback.

Research discussed in the present paper also involved two intermediate forms of discovery learning. *Moderately guided discovery* involves teacher-chosen tasks or games with modest and implicit scaffolding, such as sequentially arranged items to underscore a relation so as to prompt its implicit recognition, entering sums on a number list, and feedback on correctness only. *Minimally guided discovery* is similar to moderately guided discovery but with less implicit scaffolding. For example, instead of juxtaposing related items to underscore their connection, items are presented in no particular order.

Prior efforts

The initial software programs developed and evaluated by the authors involved minimally or moderately guided discovery learning of the add-1 rule and doubles tactics. Although recent research reviews suggest that such relatively unguided approaches are ineffective (Alfieri et al., 2011; Clark, Kirschner, & Sweller, 2012; Kirschner, Sweller, & Clark, 2006), there are three reasons to believe that a moderately, or even a minimally, guided add-1 program might be effective:

1. Although Alfieri et al. (2011) concluded that what they called unguided discovery (but perhaps more accurately fits our category of minimally guided discovery) was not an effective instructional technique, they noted that those differences were moderated by the content of instruction and the differences in outcomes based on instructional methods were less clear for mathematics.
2. Minimally guided instruction, which involved a manual number-list game and feedback regarding correctness, helped typically developing primary-grade pupils to discover and apply the add-0 rule (adding 0 to another number does not change the number) and the add-1 rule (Baroody, 1989, 1992). Although methodologically limited (untimed mental-addition task and no control group), these results indicated that a computer-assisted, minimally guided program that involved unordered practice with add-1 combinations and non-examples of the add-1 rule (add-0 or other non-add-1 combinations) might be more effective than regular classroom arithmetic instruction.
3. Hmelo-Silver, Duncan, and Chinn (2007) argued that Kirschner et al.'s (2006) category of discovery learning was extremely broad or undifferentiated and that—in contrast to unscaffolded (unguided or minimally guided) instruction—scaffolded problem-based or inquiry-based instruction (moderately or highly guided discovery) is effective. Indeed, Alfieri et al. (2011) found that such guided discovery was more effective than other forms of instruction in promoting the learning of new content.

An evaluation of the initial software (a training experiment involving a timed mental-addition task, random assignment, and a control group) indicated that both the minimally and moderately guided discovery programs were more efficacious than regular classroom instruction in helping kindergarten or first-grade children with a risk factor for academic difficulties learn and achieve fluency with the add-1 rule (Baroody et al., 2013). Surprisingly, the moderately guided discovery program was not more efficacious than the minimally guided discovery program. A follow-up training experiment revealed that even a much-improved and highly guided add-1 program was not more efficacious than the minimally guided discovery program (Purpura, Baroody, Eiland, & Reid, 2012).

The results with the doubles programs stood in stark contrast with those of the add-1 program (Purpura et al., 2012). The highly guided doubles program, but not the minimally guided program, was more efficacious than regular classroom instruction in helping at-risk first graders learn and generalize the doubles tactics. However, the effect size was relatively small (Hedges' $g = .29$) and transfer was limited (i.e., evident on only 28% of unpracticed doubles).

Motivation for the present study

The overarching goal of the present study was to evaluate the efficacy of the highly guided add-1 and doubles programs, which were again refined after the Purpura et al. (2012) evaluation. Discussed in turn are an overview of program improvements and the research aims and hypotheses tested.

Program improvements

All the programs, including the minimally guided practice-only program, were improved by more effectively preparing children for (a) using the interventions' virtual manipulatives, (b) transitioning from concrete addition to mental addition, and (c) negotiating speeded tests of mental-arithmetic. All programs included a wider variety of computer games to maintain pupil interest than did earlier versions. Specific feedback was added (e.g., explanations for why an answer or choice was incorrect were added to feedback about correctness) and more explicit hints were provided as needed. Subtraction complements of add-1 and doubles combinations were added for contrast to better ensure that a participant responded to items thoughtfully and applied learned strategies appropriately.

Three refinements were made in the present highly guided add-1 training to enhance its effectiveness over the previous versions: (a) As providing contrasting instances or non-examples of a concept can facilitate conceptual learning (Hattikudur & Alibali, 2010; Rittle-Johnson & Star, 2007), subtraction of 1 was added as a contrast to adding 1 more. (b) The relation between adding with 1 and known number-after relations was better underscored. Specifically, a new *Clue* game was added that explicitly asked, for instance, which number-after relation would help answer an add-1 item (e.g., What would help answer 3 + 1: After 3 is 4, After 6 is 7, After 7 is 8, After 9 is 10, or No good clue?). In a similar vein, a new *Does It Help?* game was added that explicitly asked whether a specific number-after relation would help with a specific add-1 item (e.g., Does knowing the number after 6 is 7 help you answer 6 + 1 = ?). (c) Some of the programs presented feedback that juxtaposed add-1 items and their related number-after relation (e.g., positioning "3 + 1 = 4" immediately above "the number after 3 is 4").

The highly guided doubles program was improved over previous somewhat successful efforts (Baroody, Eiland, Bajwa, & Baroody, 2009; Baroody et al., 2012, 2013; Purpura et al., 2012) by replacing activities that involved connecting the doubles to skip counting (e.g., 4 + 4 is analogous to counting by fours twice) with relating the sums of the doubles to even numbers and to everyday analogies of the doubles. (Previous evaluations indicated children had trouble with relating doubles to skip counting, in no small part, because they were not familiar with most skip counts.) The *Clue* and *Does It Help?* games, which explicitly asked whether a certain real-world analogy helped answer a particular double combination, were also added.

Aims and hypotheses

A primary aim of the present research was to evaluate the efficacy of the improved, highly guided discovery learning software designed to promote deliberate and then fluent use of add-1 and doubles strategies (Hypotheses 1 and 2 below). Evaluation of efficacy in the present study was methodologically more rigorous than previous efforts (Baroody et al., 2009, 2012, 2013; Purpura et al., 2012) in several respects (e.g., random assignment within class and post-testing by testers blind to training condition) and a broader sample of grade levels to increase the external validity of our conclusions. A second primary aim of the study was to compare the effects of the highly guided discovery learning to minimally guided discovery learning (Hypothesis 3). A third primary aim was to gauge whether the efficacy of an intervention varied by grade level (Hypothesis 4).

1. The computer-based add-1 program involving highly guided discovery (add-1 training or condition) and the minimally guided supplemental practice (practice condition) will be more efficacious than regular add-1 classroom training plus training on a different reasoning strategy in fostering fluency with practiced add-1 items. However, only the add-1 training will be more successful in fostering a general add-1 rule and, thus, transfer to unpracticed add-1 items.
2. Computer-based doubles intervention involving highly guided discovery (doubles training or condition) and the minimally guided supplemental practice (practice condition) should be more efficacious than regular classroom training on the doubles (and training on a different reasoning strategy) in nurturing

Table 1
Characteristics of the participants by condition.

	Condition		
	Highly guided add-1	Highly guided doubles	Minimally guided practice
Original *n*/attrition	*n* = 34/3 moved[a]	*n* = 36/5 moved[b]	*n* = 33/2 moved[c]; 1 refused[d]
Age range	5.1–7.6	5.0–7.7	5.2–7.7
Median age	5.9	6.0	6.0
Number of boys:girls	18:13	14:17	14:16
Grade			
K	17	16	18
1	10	10	9
2	4	5	3
TEMA-3 range	60–112	62–105	67–112
Median TEMA-3	89	87.5	84
Mean (*SD*) TEMA-3	88.86 (10.44)	86.46 (11.10)	87.96 (14.44)
Free/reduced lunch eligible	21	19	22
Black/Hispanic/Multiracial	23	18	19
Family history			
Single-parent	13	14	14
Parent under 18	1	1	0
Parents w/out HS	1	1	0
ESL	2	2	1
Physical condition			
Birth complications	0	0	1
Language delay	3	2	3
Speech services	5	1	5
Behavioral condition			
ADHD	1	3	2
Aggressive	4	9	3
Passive/withdrawn	2	2	3

Note. All entries other than original *n* are for participants who completed the study.

[a] (1) School 1, grade K, 5.7 years, boy, TEMA = 73, African-American; (2) School 3, grade K, 5.6 years, boy, TEMA = 102, Asian; (3) School 3, grade 2, 7.5 years, girl, TEMA = 87, free-lunch, African-American.

[b] All from School 1. (1) Grade K, 5.0 years, girl, TEMA = 62, free-lunch, African-American; (2) Grade K, 5.9 years, girl, TEMA = 90, free-lunch, African-American; (3) Grade K, 6.0 years, boy, TEMA = 66, free-lunch, Multiracial; (4) Grade 1, 6.6 years, girl, TEMA = 87, African-American; (5) Grade 1, 6.8 years, girl, TEMA = 87, African-American.

[c] (1) School 1, grade K, 5.4 years, girl, TEMA = 82, African-American; (2) Grade K, 6.0 years, boy, TEMA = 74, free-lunch, Hispanic.

[d] School 2, Grade K, 5.3 years, girl, TEMA = 95, Hispanic.

fluency with practiced doubles. However, only the doubles training will be more successful in promoting general doubles relations and, thus, transfer to unpracticed doubles.

3. Both enhanced highly guided discovery (add-1 and doubles) programs should better promote fluency with unpracticed items, if not practiced items, than minimally guided discovery involving semi-random practice, entering a sum on a number list, and feedback regarding correctness only (practice program).
4. The impact of intervention involving highly guided discovery (add-1 and doubles programs) should increase with grade level, because developmental readiness to learn reasoning strategies should improve with age/experience.

Method

Participants

Participants were recruited from 20 Grade K-to-3 classes in three elementary schools in two local school districts serving two, medium-size, mid-western cities in the U.S. Two of the schools served large populations of children with risk factors such as low mathematics achievement or poverty; the third school served a middle-class neighborhood. Parental consent forms were returned for 190 of 320 sent out. Ten children did not participate in the study because they either moved prior to subject assignments ($n = 8$) or were unable to participate in testing due to autism ($n = 2$). Another 88 children (including all Grade-3 pupils) tested out (i.e., demonstrated fluency on add-1 and doubles sums at pretest). A total of 92 pupils were eligible for the study—that is, not fluent on more than 50% of the $n + 1$ or $1 + n$ items (mean = 9%; median = 0%) or the doubles (mean = 10%; median = 0%) at pretest. A total of 11 students did not complete the study because they either moved ($n = 10$) or refused to participate ($n = 1$). The 11 students who dropped out of the study did not score lower on the The Test of Early Mathematics Ability—Third Edition (TEMA-3; Ginsburg & Baroody, 2003) than the students who completed the study, $F(1, 91) = 2.25$, $p = .138$. Moreover, attrition was equally distributed across groups and was primarily due to factors outside of the study (e.g., family mobility). Descriptive information on participants who completed the study can be found in Table 1. Participating pupils ranged in age from 5.0 to 7.7 years old (mean = 6.1). Of these children, 48% of the children were female. The majority of children were African-American (54%; 26% Caucasian; 5% Hispanic; 15% multi-ethnic, unknown, or other race). Additionally, 77% of participants were eligible for free or reduced-price lunch.

The 11 classes in Schools 1 and 2 used *Everyday Mathematics* (UCSMP, 2005). The nine classes in School 3 aligned their mathematics lessons with the School District's goals (http://www.usd116.org/index.php/parents/cia/curriculum/), used *Math Expressions* (Fuson, 2006), and adopted supplemental materials to varying degrees. Details on each of the curricula can be found in Table 2. Both curricula include activities for both group and individual work with manipulatives and materials common to many primary classrooms. No program included instructional software. Although teachers provided computer time to play math games, given the scarcity of guided discovery software for young children, it is likely these programs focused on drill of number skills. All schools were committed to achieving the State's kindergarten to grade-2 objectives that included operations on whole numbers such as solving one- and two-step problems and performing computational procedures using addition and subtraction <www.isbe.net/ils/math/standards.htm>.

Table 2
Summary of the mental-addition training and practice provided in each of the three Grade 1 curricula used in the schools.

Curriculum	General	Add-0 combinations	Add-1 combinations	Doubles combinations
Everyday mathematics (University of Chicago School Mathematics Project (2005)) (Note: Information cited in this table appears in Teacher's Manual but not student worksheets.)	Spends weeks on memorizing the "easy facts" (e.g., Unit 2 Lesson 3, Unit 4 Lessons 11 & 12, Unit 5 Lessons 10 & 11; Unit 6, Lesson 1): $n+0$, $n+1$, doubles, and sums to 10. The 0 + 0 to 9 + 9 items are practiced often (e.g., Unit 3 Lesson 14 and Unit 4 Lesson 11 use dominoes); using a "turn-around fact" (the same numbers are being added so they have the same sum) as a memory shortcut (e.g., Unit 5 Lesson 11). Worksheets with $n+0$ in row 1, $n+1$ in row 2, and doubles along a diagonal—each shaded in a different color—assigned repeatedly.	$N+0$ is related to "the sum is always the same as the number you start with."	$N+1$ is related to the sum is always the number that comes after the number you start with. Unit 1 Lesson 1 introduces 1 more by practicing number-after and adding 1 separately or implicitly (e.g., using a number line to add 1). In Unit 3, Lesson 8, a teacher explains: "Counting up by 1s is like adding 1 to each number to get the next number."	Doubles are highlighted as the numbers along the diagonal and are always even. Skip counting, such as counting by twos (2, 4, 6, . . .) and fives (5, 10, . . .), is practiced (e.g., Unit 3 Lesson 9).
Math expressions (Fuson, 2006)	Addition is introduced in terms of "decomposition" (e.g., the "break-aparts of the number 4" = 1 + 3, 2 + 2, and 3 + 1). Attention is drawn to commuted items ("switch facts"). Counting-on is encouraged initially. Mountain Math is used to introduce missing-addend addition. Visualization of finger representations and groups of 5 is encouraged (e.g., for 5 + 3: 5 fingers plus another 3).	For the most part, the grade 1 curriculum focuses on the counting or natural numbers. The concept of zero and the numeral 1is introduced in Activity 2 of Lesson 1 in Unit 1 along with the numerals 1 and 2. The "Game Cards" 0–2 are introduced in extension activities for differentiated instruction. Identifying and writing 0–2 continues in the next lesson. The Game Cards 0–10 are not listed in further units in the materials chart on p. xviii. Adding with 0 is not introduced or practiced.	A single whole-class activity explores +1 and −1 patterns: "Exploring 1 more and 1 less" (Activity 3 in Lesson 7 of Unit 1). Children holding a number card with a numeral 1–5 are lined up in order. To under-score that each number is 1 more than its predecessor, a child steps forward at the appropriate time. After "Here's 1," "Here's 1 and 1 more" is announced, and the second child steps forward up to "Here's four and 1 more." Practice is again provided in the "Consolidation" lesson (Lesson 17).	Doubles taught in terms of the even numbers and "equal sharing" in Lesson 16 Unit 1. Defined as "two groups [with] equal shares" (p. 94); pictured as paired drawings (an even number of items split into two equal groups. The flash doubles (e.g., to 5 fingers +5 fingers = 10 fingers and later to 9 + 9 = 18), doubling a single or multiple squares on graph paper, relating doubles pattern to counting by twos and dice patterns, decomposing a double such as 8 into two 4s is covered in Lessons 1 (investigate doubles) & 2 (problem solve with doubles) of Unit 7.

Measures

The achievement and fluency tests used are described in turn. See Appendix A (online supplemental material) for additional details on the fluency measure.

General mathematics achievement test

The TEMA-3 (Ginsburg & Baroody, 2003) was administered to determine if a participant was at risk for mathematical difficulties because of low achievement. The TEMA-3 is a manually and individually administered, nationally standardized test of mathematics achievement for 3- to 8-year olds. The test measures informal and formal concepts and skills in the following domains: numbering, number-comparison, numeral literacy, combination fluency, and calculation. Cronbach's alpha for the form used (Form A) overall is .94. The form's test–retest reliability is .83 and its coefficient alphas for males, females, European, African, Hispanic, and Asian Americans are all .98 and for low mathematics achievers, .99. In terms of criterion-predictive validity, correlations between the TEMA-3 and similar measures (Diagnostic Achievement Battery, KeyMath-R/NU, Woodcock-Johnson III, and Young Children's Achievement Test) range from .54 to .91.

Fluency test

The fluency test provided the data for the dependent measures. The categories of items tested and the testing procedure are discussed in turn.

Item categories

The test of fluency included four primary categories of addition items: (a) Add-1 items practiced by the add-1 and practice groups (eight *practiced add-1 items*); (b) add-1 combinations not practiced by any group (eight *unpracticed add-1 items*); (c) doubles practiced by the doubles and practice groups (five *practiced doubles*); (d) doubles not practiced by any group (five *unpracticed doubles*). The test also included non-examples of the add-1 rule practiced by only the add-1 group (2 items); two sets of non-examples of the add-1 rule practiced by the add-1 and practice groups (four *practiced add-0* and three *practiced* $n - 1$ items); non-examples of the add-1 rule or the doubles relations practiced by the add-1 and doubles groups (two items); and non-examples of the doubles practiced by the doubles and practice groups (the practiced subtraction complements to the doubles or *practiced* $2n - n$: five items). The test also included two sets of items not practiced by any group: four *unpracticed add-0* (0 + 5, 0 + 9, 6 + 0, and 7 + 0) and two commuted versions of non-examples of the add-1 rule. The specific items in each category the Cronbach's alphas for the pretest and posttest, and which group or groups, if any, practiced each category during their training are delineated in Table A-1 in Appendix A of online supplemental material.

Test procedures. Fluency testing was done in the context of computer games developed for the project. The tester encouraged a child to respond as quickly and accurately as possible without using fingers or other objects. For the *Race Car Game* (see Fig. A-1, Appendix A, online supplemental material), for example, a tester explained to the child, "We are going to play a game where we pretend you are driving a race car. In order to drive your car, you will need to keep both of your hands on the car's steering wheel at all times. [The tester then encouraged the child to grip a steering wheel tightly with both hands so as to discourage finger counting or at least make it difficult or obvious.] Your success driving your car and doing well in the race is determined by answering addition problems accurately and quickly. If you know the answer, tell me as quickly as you can. If you are not sure of the answer, make a good guess as quickly as you can." Children were not provided objects or encouraged to use objects during the fluency testing. This was done to determine if they had fluent mental addition.

Each of two testing session consisted of a test set of 12 items, computer reward game, a second test set of 12 items, and a final reward game. The items were presented in a partially random order with the constraints that two items with the same addends or sum were not presented one after another, commuted items were not presented in the same session, and the types of items were evenly distributed across the four sets.

Fluency scoring

Rationale

The goal of the research was to evaluate the efficacy of two experimental instructional programs for helping children move from Phase 1 of meaningful learning of the basic number combinations (reliance on counting) to Phase 2 (deliberate reasoning), and ultimately to Phase 3 (fluent reasoning or recall). Therefore, scoring needed to take into account how children arrived at an answer as well as efficiency (the accuracy and speed of the answer).

- *Need for evaluating strategy*. If a child in the doubles condition, for example, counted quickly to determine the correct sum of practiced and unpracticed doubles items on the delayed posttest (instead of using doubles relations), then clearly the doubles training was unsuccessful. Simply scoring the child as correct on the basis of efficiency alone results in a false positive and overestimating the impact of the training. For this reason, distinguishing among counting strategies (a Phase 1 level performance), deliberate (slow) reasoning strategies (Phase 2 level performance), and fast reasoning strategies (a Phase 3 level performance) was essential.
- *Need for scoring in context to detect response biases.* Simply scoring a child's response as correct or incorrect does not take into account whether knowledge of a strategy was applied appropriately or selectively—an important criterion for fluency. Research indicates that some children resort to response biases (to unthinkingly provide some answer), over-apply a strategy, or overgeneralize a pattern on a mental-arithmetic test. For example, some mental-addition novices consistently state the number after the larger addend, whether the addition item involves one or not (e.g., answer "six" for 0 + 5, 3 + 5, 5 + 0, and 5 + 2, as well as for 1 + 5 and 5 + 1; Baroody, 1989, 1992; Baroody, Purpura, Reid, & Eiland, 2011; Dowker, 1997, 2003). When children respond unselectively in such a manner, they accidently get the sum of 1 + 5 and 5 + 1 correct. Scoring such a response as correct results in a false positive—incorrectly indicates that the child has learned the add-1 rule or (for answers of <3 s) used the rule fluently (i.e., efficiently and appropriately/selectively).

Data collection

In order to gauge fluency as accurately as possible, testers gathered data on response accuracy, response time, and strategy. The procedure for recording accuracy and response time data is described in Cells b and c of Fig. A-1, Appendix A of the online supplemental material. Testers identified whether a child used a counting strategy, a reasoning strategy, or an undetermined strategy. This determination was based on evidence of counting objects (e.g., fingers), verbally citing (a portion of) the counting sequence out loud (even if whispered), or subvocal counting accompanied by successive movement of a finger or the eyes. A reasoning strategy was scored if a child spontaneously exhibited evidence of using deductive reasoning (e.g., for 7 + 1: "After 7 is 8"). Note that, although it is not possible to distinguish between using the add-1 rule and abstract counting-on, research by the first author (Baroody, 1995) and Bråten (1996) indicate that the former develops prior to and provides a basis for inventing the latter. Specifically, within days of discovering the add-1 rule, children begin to count-on two (e.g., "7 + 2, one more is eight, two more is nine") and then three (e.g., "7 + 3, one more is eight, two more is nine, three more is ten"). The add-1 rule is not necessary for learning the less sophisticated concrete counting-on strategy (e.g., for 5 + 2: put up two fingers, state the cardinal value of the larger addend "five" and count from there on the fingers "six" [pointing to the first finger], *seven* [pointing to the second finger]). However, children would have to represent an addend ahead of time. This takes time and would probably be seen by a trained observer.

Criteria for defining fluency

A *fluent* (Phase 3) response was defined operationally as the efficient and appropriate application of a rule/strategy: the correct answer via a reasoning or an undetermined (but not a counting) strategy in less than 3 s and *not* due to a response bias. The procedure for determining a response bias is described and illustrated in Table A-2 and Table A-3, respectively, in Appendix A of the online supplemental material. A child was deemed as using a response bias during a session (over two sets) if a potential mechanical strategy was over-applied to inappropriate cases for at least 50% of the errors made in these cases, applied to at least 50% of the trials where it could lead to an apparently correct answer, and used on at least 50% of all trials. For example, non-selective use of the add-1 rule was inferred for Session 2 for an add-1 participant who stated the number after an addend for 7 of the 13 mental-arithmetic errors, 7 of 7 add-1 items, and 14 of all 24 trials. (The child was apparently selectively applying the add-0 rule, consistently correct on the four items involving 0, and did not over-apply the add-0 rule to items not involving 0.) A research assistant (the third author) and a computer program independently assessed whether each participant used a response bias during a session. This straightforward scoring procedure yielded 98.7% and 100% inter-rater agreement on use of response bias that might produce a false positive and fluent versus non-fluent scores, respectively, on all items for all 81 participants over both tests. The dependent measure for both practiced and unpracticed combination categories was *fluency rate*: the mean proportion of a child's responses to the items of a combination category scored as fluent. The fluency rate with unpracticed items served to gauge transfer or adaptive application and, thus, *fluency with a general rule or strategy*.

Interventions

The experimental interventions supplemented regular classroom mathematics instruction and involved two 30-min sessions per week. The programs each involved five stages of computer-based training, each building on the previous stage (see Table 3). Stages I and II served as a preparation for the primary interventions, were identical for all participants, and were completed before the fluency pretest and random assignment (see Appendix B, online supplemental material, for details and sample screenshots). Stages III–V served as the primary training and differed by training condition in terms of combination family (adding with 1 or doubles) or experimental condition (highly guided or minimally guided instruction). See Appendices C–E, online supplemental material, for details and examples.

Preparatory training.

The two-stage preparatory training was designed to ensure that children had the computer skills (e.g., how to use the virtual manipulative and enter responses) needed for the testing and primary training and to provide the developmental basis (e.g., mastery of the prerequisite knowledge) for mental addition and discovery of add-1 and doubles relations. The specific number and operations goals are illustrated in the Log for Session 1 of the preparatory training (see Appendix B, online supplemental material).

Stage I. Stage I training took place over seven 30-min sessions. The aim of Stage I was to support use of counting strategies (Phase 1 in the meaningful memorization of combinations) and ensure recognition and understanding of the formal symbolism for addition and subtraction (e.g., 7 + 1 or 8 – 2) by connecting it to concrete or meaningful situations and their own informal solutions. For example, items presented as meaningful word problems *and* symbolic expressions to help children connect their symbolic word-based knowledge to formal written symbolic knowledge. Children were encouraged to solve problems in any way they wish, including informal counting-based strategies. To this end, virtual manipulatives, such as 10 frames and number sticks, were presented as an option.

During each of the seven sessions, children completed 3 sets of tasks (i.e., 1A, 1B, 1C, . . ., 7A, 7B, 7C). Set 1A served as a vehicle for learning how to navigate the program (e.g., use the mouse). Set 1B and 2A to 7A introduced virtual manipulatives (e.g., record a score using a ten frames and dots). Set 2B to 7B involved solving word problems (relating expressions or equations to a concrete model). Set 1C to 7C focused on relating part-whole terminology to equations and composition and decomposition, which underlie a number of reasoning strategies.

Stage II. Stage II training took place over 8 sessions (sets 8–15). The aim of Stage II was to serve as a developmental bridge between using Phase 1 strategies (i.e., informal counting-based strategies with objects such as fingers or a ten frame) promoted by Stage I and Phases 2 and 3 (i.e., using mental-arithmetic strategies involving reasoning or retrieval). Essentially, the goal of Stage II was to help children identify and define a "good" or "smart guess." Numerical estimation (approximating the size of a single collection) was introduced first in sets 8 and 9. Arithmetic estimation (approximating the size of sums and differences) was introduced in sets 10–12. The stage begins with visually estimating the *number* of carrots or frogs. (About how many carrots or frogs did you see?) This provides a basis for estimating the answers to *addition and subtraction* problems that come next (see Figs. B-1 and B-2, respectively, in Appendix B, online supplemental material).

Primary training

The add-1 and doubles programs were tailored to promote *highly guided* discovery and practice of a particular reasoning strategy. In an effort to move beyond passive learning activities (Chi, 2009) and consistent with the recommendation of the NMAP (2008), Stage III and often Stage IV training involved sequentially arranging problems to highlight a relation and where that relation was applicable. For example, for the add-1 training, answering a number-after-n question (e.g., "What number comes after 3 when we count?") was immediately followed by answering a related $n + 1$ item (e.g., "3 + 1 = ?"). The $1 + n$ counterpart was posed next to prompt recognition of additive commutativity and the applicability of the add-1 rule to $1 + n$ items. An add-0 item and an $n + m$ item (where n *and* $m > 1$) served as non-examples of the add-1 rule to discourage over-generalizing the rule (i.e., promote its appropriate use). Similarly, in the doubles training, the doubles were connected to fair sharing and even numbers. Some instruction and feedback encouraged children to think about using what they knew to shortcut computational effort. The training for the practice group involved *minimally guided discovery of the add-1 rule and doubles relations*. It consisted of engaging in similar activities (e.g., games involving a number list), but the order of items was haphazard,

Table 3
Five stages of computer-based mental-arithmetic training stages I & II = common preparatory training; Stages III–V = condition-specific experimental training.

Stage number and focus	Description
Stage I: Concrete training (7 sessions; ~3.5 weeks)	**Aim**: *Phase 1* (counting-based) and computer/virtual manipulatives training. Ensure understanding of symbolic addition and subtraction expressions (e.g., 7 + 1) or equations (e.g., 8 − 2 = 6) by connecting them to solving meaningful word problems. **Plan**: For each of 7 sessions, there were three sets. Set 1A served a vehicle for learning how to navigate the program (e.g., use the mouse). Sets 1B and 2A to 7A introduced virtual manipulatives (e.g., record a score using a ten frames and dots). Sets 2B to 7B involved solving word problems (relating expressions/equations to a concrete model). Child free to choose strategy, including the use of virtual manipulatives, such as ten frames and number sticks, encouraged. Untimed. Set 1C to 7C focused on relating part-whole terminology to equations and composition and decomposition.
Stage II: Estimation training (8 sessions; ~4 weeks)	**Aim**: Serve as a bridge between Phase 1 and mental-addition (Phases 2 and 3). Help define a good or "SMART GUESS." **Plan**: Numerical estimation (approximating the size of a single collection) was introduced first in Sets 8 and 9; arithmetic estimation (gauging the size of sums and differences), in sets 10–12. Emphasized was a range of reasonable answers on a number list. Untimed/timed.
Stage III: Strategy training (8 sessions; ~4 weeks)	**Aim**: *Phase 2* (overt reasoning) training. Help a child discover the relations that underlie a reasoning strategy and thus understand and effectively use a reasoning strategy: Add-1: add-1 items followed related number-after relation; Doubles: doubles followed everyday visual analogies or the use of a fair-sharing analogy to identify even numbers on a number list. Practice: related items to those with the same answer (e.g., commuted item). **Plan**: Items presented concretely (number list and ten frames) and symbolically. Child encouraged to use a strategy of his/her choice. No time limit was set. For incorrect responses, child directed to redo using virtual manipulatives and counting. Related items immediately follow one another; feedback juxtaposed equations.
Stage IV: Strategy practice (8 sessions; ~4 weeks)	**Aim**: Promote *Phase 3*—efficient use of reasoning strategies. **Plan**: Items presented symbolically. Child encouraged to respond mentally and quickly ("make a smart guess as fast as you can"). Generous (6 s) time limit. For incorrect responses, feedback encouraged child to use the relationship they are being trained in to solve problems in a timely fashion. Concrete solutions (number lists and ten frames) used only as a last resort. Related items juxtaposed or immediately follow one another only some of the time.
Stage V: Strategy fluency (8 sessions; ~4 weeks)	**Aim**: Cement efficient use of reasoning strategies (Phase 3). **Plan**: Items presented symbolically. Child encouraged to respond mentally and quickly ("make a smart guess as fast as you can"). Stringent (3 s) time limit. For incorrect responses, child given a second-chance guess, no hints, no manipulates. Feedback emphasized the advantages of a fast smart guess; disadvantages of counting. Items presented in a semi-random order (related items not juxtaposed).

connections to prior knowledge were not made, and feedback provided consisted of whether a response was correct or not.

Research design and procedures

A training experiment involving the assessment of multiple baselines was used. During the preparatory training, participating students completed the TEMA-3 to gauge general mathematics achievement ability. After the preparatory training, the fluency pretest was administered to assess baseline fluency with add-1 and doubles sums. Eligible participants were then randomly assigned *within* class to one of three primary/experimental training conditions to control for class and school confounds. All participants were re-tested on the fluency test two weeks after the completion of the primary training. This delayed fluency posttest served to gauge retention of practiced combinations and transfer to unpracticed combinations. All project training and testing was conducted on a one-to-one basis at project computer stations in a hallway outside a child's classroom or in a room dedicated to the project. Pull-outs occurred in non-literacy time blocks, including mathematics instruction and playtime—as per the request of the districts.

Two different active-control groups were used. (a) As the add-1 and doubles training targeted different combination families, each served as a comparison group for testing the efficacy of the other. As the doubles group did not receive experimental add-1 training or supplemental add-1 practice, for example, it served to control for the effects of outside (e.g., classroom) training and practice involving add-1 combinations. (b) As the practice group practiced add-1 and doubles items as often as the add-1 and doubles groups, respectively, it served as a control for the effect of supplemental practice alone. This active-control also provided a basis for comparing the relative impact of highly and minimally guided discovery.

Analyses

Efficacy was evaluated by comparing the mean posttest fluency rate of each group on targeted practiced and unpracticed combinations in two ways. (a) One way was with ANCOVAs, using pretest fluency rate and age as the covariates. Effects of treatment were tested using one-tailed significance values given the directional nature of the contrasts (e.g., for the add-1 analyses the structured add-1 group > unstructured practice > control (doubles) group and for the structured doubles analyses the doubles group > unstructured practice group > control (add-1) group. As children were randomly assigned to condition within school and within class, these variables were not expected to confound condition. Nevertheless, all analyses were also conducted with school and class, as well as grade level (which reasonably might affect an intervention's impact), as random-effect covariates. The Benjamini–Hochberg correction was applied to correct for Type I error due to multiple comparisons. (b) A second way efficacy was evaluated was by calculating an effect size (Hedges' g) for all contrasts using the mean posttest fluency rates adjusted for age and pretest fluency rate. This was done for three reasons. One is that

Table 4
Pretest mean proportion fluent and delayed posttest adjusted mean proportion fluent and (in parenthesis) standard deviation by condition for practiced and unpracticed add-1 and doubles combinations.

Condition	Add-1 combinations				Doubles combinations			
	Practiced		Unpracticed		Practiced		Unpracticed	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Add-1 group	.09 (.13)	.47 (.32)	.06 (.12)	.36 (.32)	.15 (.22)	.27 (.28)	.06 (.12)	.09 (.15)
Practice group	.12 (.20)	.54/.55[a] (.36)	.07 (.12)	.36/.37[a] (.30)	.15 (.21)	.59/.60[a] (.31)	.07 (.14)	.11 (.16)
Doubles group	.11 (.16)	.25/26[b] (.31)	.07 (.12)	.27/.28[b] (.28)	.15 (.24)	.60/.61[b] (.30)	.06 (.14)	.17 (.21)

[a] The first entry in each case is the adjusted mean for the add-1 versus practice comparison, and the second, for the practiced versus doubles comparison.
[b] The first entry in each case is the adjusted mean for the add-1 versus doubles comparison and the second, for the practice versus doubles comparison.

Lipsey et al. (2012, p. 3) noted that a significance level does not bear a necessary relationship to "practical significance or even to the statistical magnitude of the effect. Statistical significance . . . is heavily influenced by the sample size, the within samples variance on the outcome variable, the covariates included in the analysis, and the type of statistical test applied. None of [which] is related in any way to the magnitude or importance of the effect." A second is that Wilkinson and the APA Task Force on Statistical Inference (1999) recommended such a statistic when the power of a study is limited and a significance test may not detect a real effect. A third reason is that the Institute of Education Science (IES) supported the research, and the Institute's What Works Clearinghouse (WWC) guidelines (IES, 2014) set a $g \geq .25$ as the criterion for a *substantively important practice*.

Results

Detailed in Table 4 are the mean pretest fluency rate, the adjusted posttest mean fluency rate, and the standard deviations of practiced and unpracticed add-1 and doubles combinations by combination type and condition. Preliminary analyses revealed that the groups did not differ significantly in mean fluency rates at pretest on practiced add-1 items ($F[2, 78] = 0.22$, $p = .803$), unpracticed add-1 items ($F[2, 78] = 0.13$, $p = .883$), practiced doubles ($F[2, 78] = 0.01$, $p = .992$), unpracticed doubles ($F[2, 78] = 0.1$, $p = .990$), or on the TEMA-3 ($F[2, 78] = 0.28$, $p = .766$). The data for response biases with the potential to create false positives are summarized in Appendix F (online supplemental material). School was not a significant factor for any of the primary analyses (i.e., those that bear on the hypotheses) or secondary analyses (for items other than add-1 or doubles). Teacher was found to be a significant predictor for three analyses, but nevertheless did not meaningfully change the results. Thus, the results are all reported without school and teacher variables included. Grade level was found to be a significant predictor for one primary analysis, but this did not change the primary effect and thus grade is not reported in the analyses with one exception. A significant grade × condition interaction effect was found for the unpracticed doubles analysis that compared the doubles and practice groups. This interaction is discussed in the Hypothesis 4 subsection of the Primary Analyses. All significant results remained so after applying the Benjamini-Hochberg correction.

Primary analyses

Hypothesis 1. Efficacy of the add-1 programs.

For practiced add-1 combinations, planned contrasts revealed that, at the delayed posttest, the add-1 and practice groups significantly outperformed the doubles group, which did not receive experimental training with these items, $F(1, 50) = 7.71$, $p = .004$, Hedges' $g = 0.70$ and $F(1, 49) = 10.05$, $p = .002$, Hedges' $g = 0.85$, respectively. More importantly, for unpracticed add-1 items, although a significant difference was not found between the add-1 and the doubles groups ($F[1, 50] = 1.30$, $p = .130$, Hedges' $g = .28$), the effect size for this comparison meets the criterion for substantively important practice set by the WWC guidelines (IES, 2014). The same was true for the practice group versus doubles group comparison, $F[1, 49] = 1.05$, $p = .156$, Hedges' $g = .27$.

The previous analyses excluded false positives that might be due a response bias (see Table F-1 in Appendix F of the online supplemental material). Although response biases of all types decreased from fluency pretest to posttest, the number of children who resorted to a number-after-an addend or a number-after-larger-addend response bias increased. For instance, nine participants across the three conditions consistently stated an addend or the larger addend in one or both sessions on the pretest, but only four children from two (the practice and doubles) conditions did so on the posttest. In contrast, on the pretest, three, one, and one participant(s) in the add-1, doubles, and practice group met the criteria for a number-after or a number-after-larger-addend response bias. On the posttest, five, zero, and two participant(s) in the add-1, doubles, and practice group met the criteria for such response biases.

Hypothesis 2. Efficacy of the doubles programs.

For practiced doubles, planned contrasts revealed that both the doubles and practice groups significantly outperformed the add-1 group, which did not receive experimental training with these combinations, $F(1, 50) = 23.82$, $p < .001$, Hedges' $g = 1.14$ and $F(1, 51) = 22.34$, $p < .001$, Hedges' $g = 1.10$, respectively. For unpracticed doubles, the doubles group—unlike that in the Purpura et al. (2012) study—significantly outperformed the add-1 group, $F(1, 50) = 4.30$, $p = .022$, Hedges' $g = 0.43$. The practice group did not do so, $F(1, 51) = 0.37$, $p < .273$, Hedges' $g = 0.13$. Perhaps more importantly, the effect size for the doubles group (but not the practice group) is indicative of substantively important practice according to the WWC guidelines (IES, 2014).

Hypothesis 3. Guided programs should be more effective than unguided practice.

Despite the authors' best and protracted (7-year) effort to make the guided add-1 training as effective as possible, no significant differences were found between the add-1 and practice groups for the practiced add-1 items ($F[1, 51] = 0.56$, $p = .531$, Hedges' $g = -.19$) or the unpracticed add-1 combinations ($F[1, 51] = 0.00$, $p = .480$, Hedges' $g = 0.02$). As expected, the doubles group did not outperform the practice group on the practiced doubles ($F[1, 49] = 0.01$, $p = .470$, Hedges' $g = 0.02$) but did so with the unpracticed doubles at a marginally significant level ($F[1, 49] = 2.24$, $p = .071$, Hedges' $g = 0.33$). The effect size difference for the analysis is indicative of substantively important practice according to the WWC guidelines (IES, 2014).

Hypothesis 4. Impact of guided programs should increase with grade level.

When grade level was included as a random effects covariate, there was a significant grade (K, 1, 2) × condition (doubles, practice)

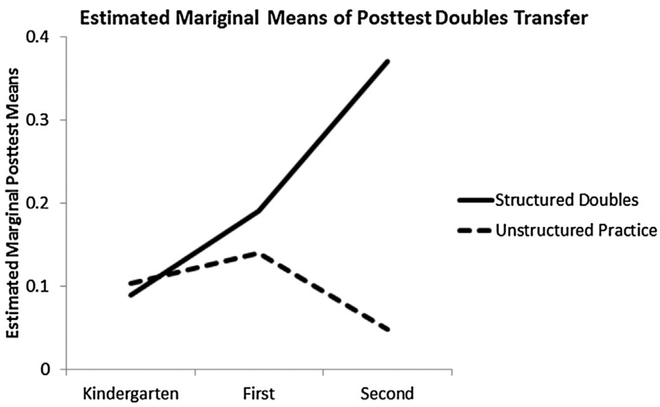

Fig. 1. The adjusted posttest means for the doubles and practice groups by grade for unpracticed doubles. The low adjusted posttest score by the practice group in grade 2 occurred because a couple participants in that group score lower on the doubles unpracticed items at posttest than they did at pretest. This had an unduly large impact because there were relatively few grade 2 participants. That is, in practical terms, it is reasonable to conclude that participants in the practice group at all grade levels made little, if any, progress in transferring doubles knowledge.

interaction for the unpracticed doubles (but not the practiced or unpracticed add-1 items or the practiced doubles) at the delayed posttest, $F(2, 45) = 3.71$, $p = .032$. As can be seen in Fig. 1, whereas the practice group exhibited only minimal transfer regardless of grade level, the doubles group exhibited an increase in their doubles transfer fluency as grade level increased. Specifically, there was no meaningful difference between participants between the groups at Grade K (Hedges' $g = -0.10$), a marginal difference at Grade 1 (Hedges' $g = 0.25$), and a large difference at Grade 2 (Hedges' $g = 1.11$). Although not statistically significant (F[2, 45] = 1.10, $p = .342$), the grade × condition for the doubles versus add-1 comparison followed the same pattern with small, small, and medium meaningful differences between participants between the groups at Grade K (Hedges' $g = 0.26$), Grade 1 (Hedges' $g = 0.42$), and Grade 2 (Hedges' $g = 0.64$), respectively.

Secondary analyses

The data for the add-0 and subtraction combinations are summarized in Table G-1 in Appendix G (online supplemental material). The add-0 family was by far the least difficult combination family for the participants. Participants were fluent on about a third of the practiced and unpracticed add-0 items at pretest—double or more the fluency with other families. At delayed posttest, participants in the add-1 intervention groups were fluent on at least 80% and 75% of the practiced and unpracticed add-0 combinations, respectively. Even the comparison (doubles) group had achieved fluency on about 67% of these combinations at posttest. The add-0 rule appears to be highly salient—even more so than the add-1 rule. This rule should be a focal point of grade-1 or even grade-K instruction, where it can serve as a useful contrast for learning the add-1 rule, whether related add-0 and add-1 items are practiced sequentially or haphazardly during a session.

Discussion

Primary aims: efficacy of the computer-based interventions

Efficacy of the add-1 program

Hypothesis 1 was partially confirmed. As predicted both highly guided add-1 training and the minimally guided supplemental practice were efficacious in fostering fluency with practiced add-1 combinations. However, contrary to Hypothesis 1, both guided add-1 *and* the unguided practice successfully promoted transfer to unpracticed add-1 combinations—results that indicate that both were efficacious in fostering fluency with a general add-1 rule.

Efficacy of the doubles program

Hypothesis 2 was confirmed. Although both the guided doubles training and the minimally guided supplemental practice were efficacious in fostering fluency with practiced doubles, more importantly the former, but not the latter, was successful in promoting transfer to unpracticed doubles. The fact that the effect size was 50% greater than the effect size reported in the Purpura et al. (2012) study also indicates that the revision to the doubles program had an impact.

The relative impact of guided versus unguided discovery and grade level

Superiority of guided instruction

Hypothesis 3 was corroborated for the doubles, but not the add-1, program. Consistent with previous findings (Baroody et al., 2013; Purpura et al., 2012) and despite the authors' 7-year effort to make the guided add-1 training as effective as possible, this program was not more efficacious than unguided practice in promoting fluency with practiced add-1 items or transfer to unpracticed add-1 combinations. In contrast, although the guided doubles program was not more efficacious than unguided practice in promoting fluency with the practiced doubles, it was more efficacious in promoting transfer to unpracticed doubles. Apparently, guided training is important for helping children learn and generalize doubles relations, but relatively unguided supplemental practice is sufficient for learning a general add-1 rule.

Why the value of discovery learning might vary with combination family. Two reasons might plausibly account for why highly guided discovery was more effective than minimally guided instruction in teaching the doubles relations but not the add-1 rule.

Reason 1—the visual analogy of the number list. In addition to the fact that both the add-1 and practice programs involved supplemental practice with adding with 1, both programs also required a participant to enter the sums of such items on a number list and used a number list to provide visual feedback. These visual analogies directly embody the add-1 rule: Adding 1 to a number (moving 1 cell more to the right on a number list) has as its sum the next number after the number in the counting sequence (i.e., results in landing on the cell of the next larger number on the number list). In a sense, then, even the semi-random drill of the practice training was minimally structured or guided in two senses: (a) adding 1 was targeted for practice and (b) the visual analogy of number list implicitly modeled the add-1 rule. Future research needs to evaluate whether the supplemental practice, the number list analogy, or both are critical for promoting discovery of the add-1 rule by comparing negligibly guided add-1 training (semi-random targeted practice without the visual aid of a number list) with minimally guided add-1 training (semi-random targeted practice with the visual aid of a number list).

Although the practice condition also included activities that involved entering answers to the doubles items on a number list, there is no direct relation between this visual representation and doubles relations. Thus, such a representation was probably not helpful in itself, and the practice condition might better be considered negligibly guided (targeted practice only) discovery, which appears to be insufficient to induce the non-salient doubles relations. In contrast, the highly guided discovery learning of the doubles program included activities where a child had to determine if a certain number of cookies could be shared fairly and was an even number or could not be fairly shared and was an odd number. As the child identified even numbers, these numbers were highlighted in green, which underscored that every other number, starting with two, is even. After the child determined the sum of a double, the expression settled atop the number list at its sum (e.g., 4 + 4 settled atop the cell 8 highlighted in green on a number list). Connecting the doubles to this enhanced visual representation of the number line may have helped children understand that the sum of a double must be even. Using this knowledge, knowledge of "easy doubles" such as 5 + 5 = 10, and knowledge that 6 is the next bigger number after five in the counting sequence, a child could deduce that the unknown sum of 6 + 6 must be the next even number after 10. However, it remains for future research to determine if this connection to the even numbers in the doubles training, the real-world analogies, or both contributed to the success of the program.

Reason 2—salience and fluency with component knowledge. The difference in the ease of learning the add-1 rule and doubles tactics may also be due to two, inter-related, factors. One factor is the salience of the concepts (patterns or relations) underlying the meaningful learning of a strategy. Salience is determined by the complexity of a regularity itself and the prior conceptual, factual, and procedural knowledge required to induce and assimilate the pattern or relation. The second factor is prior fluency with component skills needed to execute the reasoning strategy and the complexity of these skills. Primary-age children can learn and apply the number-after rule for adding 1 via highly *or* minimally guided instruction because the regularity underlying the rule is highly salient for them and they are already highly fluent with the component skills for implementing the rule.

The connection between number-after relations is relatively straightforward. After using a counting strategy to determine that the sum of 7 + 1 is 8, for instance, a child might recall the well-known relation that *eight* comes (immediately) after *seven* in the counting sequence. Several such experiences may lead the child to conclude that the sum of any number n and one is the number after n. This is consistent with the active-memory view that practice is not merely a vehicle for strengthening an existing factual association but an opportunity to enrich extant memory by actively creating new memories (Nader & Hardt, 2009). The insight into how addition of one is connected to the counting sequence and the implementation of the resulting number-after rule for adding with 1 are facilitated by the fact that pupils just beginning school are typically well-versed in the prerequisite knowledge needed for discovery of the add-1 rule and its fluent application. Preschoolers usually discover the increasing magnitude principle—that a number further along in the counting sequence represents larger collection than numbers before it—before they turn five years of age (Sarnecka & Carey, 2008; Schaeffer, Eggleston, & Scott, 1974). Moreover, children are so familiar with the counting sequence, they can readily recall and apply number-after relations to new tasks, such as mentally adding with 1 (Fuson, 1988). As Kirschner et al. (2006) noted, although guided instruction is generally more effective than unguided instruction, this advantage recedes "when learners have sufficiently high prior knowledge to provide 'internal' guidance" (p. 75). In contrast, the doubles strategies may be more difficult to learn, even with guided instruction, because the doubles connection to the even numbers, everyday analogies, and skip counting (e.g., recognizing that the 4 + 4 can be determined by skip counting "four, eight") may not be salient, in part, because primary-age pupils are unfamiliar or non-fluent with the prerequisite knowledge.

Educational implications. Given the relative ease of discovering the add-1 rule and its importance for fluency with other reasoning strategies such as the near doubles, it makes sense to establish fluency with add-1 combinations as a Grade 1 or even Grade K goal. In contrast, given the relative difficulty of discovering doubles relations, it may make sense to postpone the goal of achieving *fluency* with all the basic doubles until the end of Grade 2 or perhaps even later. This is not to say that meaningful learning of doubles relations, such as relating them to everyday analogies and the even numbers, should not begin as early as Grade 1 or even Kindergarten.

Increasing effectiveness with grade level

Hypothesis 4 was, for the most part, not supported. The one exception was transfer of the doubles training to unpracticed doubles—arguably the most difficult of the main goals of the experimental programs. This was achieved despite the modest number of Grade 2 participants. Although further research is needed to examine the issue, the significant grade x condition interaction for the unpracticed doubles further supports the conclusion that the goal of applying the relatively non-salient doubles relations fluently might better be postponed until Grade 2.

Other educational and methodological implications

The overall results have implications regarding (a) the role of practice frequency in meaningful memorization of basic combinations, (b) concurrent practice of related aspects of knowledge, (c) the value of guided discovery learning, (d) the need for computer-based mental-addition training of at-risk children, and (e) the need to consider mental-arithmetic response biases.

The role of practice frequency

The overall results indicate that frequency may not be the most important factor in the meaningful memorization of basic combinations (cf. Ashcraft, 1992). The positive results for the two highly guided interventions with both practiced and unpracticed combinations and those for the minimally guided-practice training with practiced combinations were achieved *despite* only 27 or 28 repetitions for each of the items practiced by a child—substantially less practice than thousands of repetitions per item necessary to achieve memorization (by rote) of these facts specified by earlier models and computer simulations of arithmetic learning (Shrager & Siegler, 1998; Siegler & Jenkins, 1989). A possible implication is that a focus on mathematical regularities (patterns and relations) and (analogical, inductive, and deductive) reasoning may be more effective than unstructured and meaningless drill in promoting fluency with the basic sums. Another possible implication is that computer-based games may be an engaging way to practice basic combinations and that high interest or motivation may be as important as practice frequency, if not more so.

The role of concurrent practice

The transfer produced by the highly and minimally guided add-1 and the highly guided doubles programs on targeted families is consistent with Nader and Hardt's (2009) constructive memory hypothesis: Each time a known fact or relation is recalled, new information can be connected to it—enriching or even transforming existing knowledge and increasing an individual's ability to recall it. For example, recalling that the number after five is six while solving 5 + 1 = ? may help children to recognize the connection between adding 1 and known number-after relations and construct a general add-1 rule. Moreover, although not the case with the highly salient add-1 rule, structured practice that requires recalling existing knowledge immediately followed by practice with novel information would seem especially likely to promote the integration of extant knowledge with new information in cases where a relation is less obvious to children. For example, practice with the analogy of "5 fingers on left hand & 5 fingers on right hand make 10 fingers in all" immediately before determining the sum of 5 + 5 = ? may have primed participants to recall the everyday analogy and to link this informal and formal knowledge of 5 + 5 = 10 in memory.

The constructive memory process described in the previous paragraph fills the following important theoretical gap identified by Siegler and Ramani (2009): "Future models of arithmetic [might] benefit from including retrieval structures or other mechanisms that embody numerical magnitude representations" (p. 556). Specifically, the add-1 rule may be what connects representations of counting, numerical magnitude, and addition. Recalling that the number after five is six while solving 5 + 1 = ? may help children to construct or strengthen the successor principle: Each counting number is exactly one more than its predecessor. For this and other reasons previously cited (e.g., providing a basis for more advanced mental arithmetic), learning the add-1 rule should be a primary goal of Grade 1 or perhaps even Grade K instruction.

The value of discovery learning

The increase in fluency with practiced and unpracticed items produced by the interventions, particularly by the doubles training, provides additional supporting evidence for what Alfieri et al. (2011) called "the generation effect" with a genuine school (ecologically valid) task. They defined a *generation effect* as the enhancement of learning and retention when learners are permitted to construct their own knowledge in some way, such as generating their own generalization.

The need for computer-based intervention

Disquieting is that, with regular classroom instruction and practice, the doubles participants—almost half of whom were in Grade 1 or 2—were fluent on less than 30% and 5% of the relatively easy add-1 and minus-1 combinations at the *end* of the school year (see also Henry & Brown, 2008). The same was true for the add-1 participants with doubles and subtraction of doubles complements. Such low mental-arithmetic proficiency does not bode well for mathematics achievement later in school. As teachers in classes with many children with a risk factor for academic difficulties are already stretched to the limit, there is a need for stand-alone software that helps such children achieve mental-arithmetic proficiency.

Although little software exists to help young children discover relations or invent reasoning strategies, such software might be particularly valuable for intervention with at-risk children (Butterworth, Sashank, & Laurillard, 2011; Kucian et al., 2011; Rasanen, Salminen, Wilson, Aunio, & Dehaene, 2009). Well-designed computer programs and properly chosen computer games can provide effective instruction and practice even for young children (Clements & Sarama, 2012; NRC, 2009; Sarama & Clements, 2009). A program can provide the scaffolding for guided discovery learning that most teachers cannot. Specifically, it can underscore relations, such as the connection between number-after relations and adding 1 (see Appendix C for details, online supplemental material), a connection that may not be known, explained clearly, or emphasized by most teachers. The game context provided motivation to learn for children accustomed to being entertained by television and online/mobile/computer games. The fact that positive assent from a child had to be secured for *each* training session and that only one child was lost to attrition because this assent could not be secured is a testament to engagement.

The need to check for response biases

Although fewer children in the add-1 condition used response biases at posttest, the number of such children who consistently resorted to stating the number after an addend or the larger addend in one or both sessions increased somewhat. For example, though the number of add-1 participants who applied a state-the-addend or -larger-addend response bias in at least one session declined from four on the pretest to one on the posttest, the corresponding numbers for state-the-number-after-an-addend or state-the-larger-addend response biases increased from three to five—compared to one and zero for the doubles group and one and two for the practice group. These results are consistent with the conclusion that the add-1 training helped participants to abandon highly mechanical strategies such as state-an-addend but may have induced some children to inappropriately over-apply the number-after rule to combinations not involving one.

Educationally, these results underscore the need for instruction to highlight non-examples of a rule, such as the add-1 rule, as well as examples of the rule. Specifically, the add-1 program could be improved by building into it a checking mechanism for detecting state-the-number-after-an-addend or state-the larger-addend response biases and providing explicit feedback when the add-1 rule applies and when it does not. Primary-grade classroom teachers should also be taught to do the same. Methodologically, the results are consistent with those found by Baroody et al. (2011) that scoring in context is necessary so as to exclude false positives and not over-estimate the effects of an intervention.

Acknowledgements

The research reported herein was supported by a grant from the Institute of Education Science, U.S. Department of Education, through Grant R305A080479 "Fostering Fluency with Basic Addition and Subtraction Facts". Preparation of this manuscript also was supported, in part, by another grant from the U.S. Department of Education (R305B100017: "UIUC Postdoctoral Research Training Program in Mathematics Education"). The opinions expressed are solely those of the authors and do not necessarily reflect the position, policy, or endorsement of the Institute of Education Science or Department of Education.

Appendix A to G. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecresq.2014.09.003.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*, 1–18. http://dx.doi.org/10.1037/a0021017

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106. http://dx.doi.org/10.1016/0010-0277(92)90051-I

Baroody, A. J. (1989). Kindergartners' mental addition with single-digit combinations. *Journal for Research in Mathematics Education*, *20*, 159–172. http://dx.doi.org/10.2307/749280

Baroody, A. J. (1992). The development of kindergartners' mental-addition strategies. *Learning and Individual Differences*, *4*, 215–235. http://dx.doi.org/10.1016/1041-6080(92)90003-W

Baroody, A. J. (1995). The role of the number-after rule in the invention of computational short cuts. *Cognition and Instruction*, *13*, 189–219. http://dx.doi.org/10.1207/s1532690xci1302_2

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Mahwah, NJ: Erlbaum.

Baroody, A. J., & Coslick, R. T. (1998). *Fostering children's mathematical power: An investigative approach to K-8 mathematics instruction*. Mahwah. NJ: Erlbaum.

Baroody, A. J., Eiland, M. D., Bajwa, N. P., & Baroody, S. C. (2009, June). *Fostering primary-level at-risk children's fluency with basic addition combinations*. Poster presented at the Fourth Annual IES Research Conference, Washington, DC.

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering kindergarten children's number sense. *Cognition and Instruction*, *30*, 435–470. http://dx.doi.org/10.1080/07370008.2012.720152

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal*, *50*, 533–573. http://dx.doi.org/10.3102/0002831212473349

Baroody, A.J., Purpura, D.J., Reid, E.E., & Eiland, M.D. (2011, September). Scoring fluency with basic addition combinations in context. In A. J. Baroody (Chair), Issues in the assessment and scoring of early numeracy skills. Symposium at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Bråten, I. (1996). *Cognitive strategies in mathematics (Report No. 10)*. Oslo, Norway: Institute for Educational Research, University of Oslo.

Brownell, W. A. (1941). *Arithmetic in grades I and II: A critical summary of new and previously reported research*. Durham. NC: Duke University Press.

Butterworth, B., Sashank, V., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, *332*, 1049–1053. http://dx.doi.org/10.1126/science.1201536

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's problem solving. *Developmental Psychology*, *34*, 882–891. http://dx.doi.org/10.1037/0012-1649.34.5.882

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105. http://dx.doi.org/10.1111/j. 1756-8765.2008.01005.x

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator*, *36*, 6–11.

Clements, D. H., & Sarama, J. (2012). Learning and teaching early and elementary mathematics. In J. Carlson, & J. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (Vol. 3) (pp. 107–162). Charlotte, NC: Information Age Publishing.

Council of Chief State School Officers. (2010). *Common Core State Standards: Preparing America's students for college and career*. Retrieved from: http://www.corestandards.org/

Cowan, R. (2003). Does it all add up? Changes in children's knowledge of addition combinations, strategies, and principles. In A. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 35–74). Mahwah, NJ: Erlbaum.

Dowker, A. (2009). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development*, *24*, 401–410. http://dx.doi.org/10.1080/135467997387452

Dowker, A. D. (1997). Young children's addition estimates. *Mathematical Cognition*, *3*, 141–154. http://dx.doi.org/10.1080/135467997387452

Dowker, A. D. (2003). Young children's estimates for addition: The zone of partial knowledge and understanding. In A. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 35–74). Mahwah, NJ: Erlbaum.

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction. *Cognition*, *123*, 392–403. http://dx.doi.org/10.1016/j.cognition.2012.02.008

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 318–374). New York, NY: Macmillan.

Fuson, K. C. (2006). *Math expressions*. Boston, MA: Houghton Mifflin.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Austin, TX: Pro-Ed.

Hatano, G. (2003). Foreword. In A. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. xi–xiii). Mahwah, NJ: Erlbaum.

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: Does comparing with inequality symbols help? *Journal of Experimental Child Psychology*, *107*, 15–30. http://dx.doi.org/10.1016/j.jecp.2010.03.004

Henry, V., & Brown, R. (2008). First-grade basic facts: An investigation into teaching and learning of an accelerated, high-demand memorization standard. *Journal for Research in Mathematics Education*, *39*, 153–183. http://dx.doi.org/10.2307/30034895

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark. *Educational Psychologist*, *42*, 99–107. http://dx.doi.org/10.1080/00461520701263368

Institute of Education Sciences. (2014). *What works clearinghouse: Procedures and standards handbook (Version 3)*. http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

Jerman, M. (1970). Some strategies for solving simple multiplication combinations. *Journal for Research in Mathematics Education*, *1*, 95–128. http://dx.doi.org/10.2307/748856

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. http://dx.doi.org/10.2307/748856

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage*, *57*, 782–795. http://dx.doi.org/10.1016/j.neuroimage.2011.01.070

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., et al. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington, DC: IES National Center for Special Education Research, Institute of Education Sciences.

Murata, A. (2004). Paths to learning ten-structured understandings of teen sums: Addition solution methods of Japanese grade 1 students. *Cognition and Instruction*, *22*, 185–218. http://dx.doi.org/10.1207/s1532690xci2202_2

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews/Neuroscience*, *10*, 224–234. http://dx.doi.org/10.1038/nrn2590

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics: Standards 2000*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Reston, VA: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

Purpura, D.J., Baroody, A.J., Eiland, M.D., & Reid, E.E. (2012, September). Fostering first-graders' reasoning strategies with the most basic sums. In A. J. Baroody (Chair), Implications of three training experiments for tailoring instruction to promote the fluency of specific arithmetic families. Symposium at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Rasanen, P., Salminen, J., Wilson, A., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for numerical skills in children with low numeracy: A review and a randomized controlled trial. *Cognitive Development*, *24*, 450–472. http://dx.doi.org/10.1016/j.cogdev.2009.09.003

Rathmell, E. C. (1978). Using thinking strategies to teach basic facts. In M. N. Suydam, & R. E. Reys (Eds.), *Developing computational skills* (pp. 13–50). Reston, VA: National Council of Teachers of Mathematics.

Rittle-Johnson, B., & Star, J. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*, 561–574. http://dx.doi.org/10.1037/0022-0663.99.3.561

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*, 662–674, 10.1016/j.cognition.2008.05.007.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, *6*, 357–379. http://dx.doi.org/10.1016/0010-0285(74)90017-6

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*, 405–410. http://dx.doi.org/10.1111/1467-9280.00076

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560. http://dx.doi.org/10.1037/a0014239

Swenson, E. J. (1949). Organization and generalization as factors in learning, transfer and retroactive inhibition. In E. J. Swenson, G. L. Anderson, & C. L. Stacey (Eds.), *Learning theory in school situations. University of Minnesota Studies in Education*. Minneapolis, MN: University of Minnesota Press.

Thiele, C. L. (1938). *The contribution of generalization to the learning of addition facts*. New York, NY: Bureau of Publications, Teachers College, Columbia University.

Thornton, C. A. (1978). Emphasizing thinking strategies in basic fact instruction. *Journal for Research in Mathematics Education*, *9*, 214–227. http://dx.doi.org/10.2307/748999

Thornton, C. A. (1990). Solution strategies: Subtraction number facts. *Educational Studies in Mathematics*, *21*, 241–263. http://dx.doi.org/10.1007/BF00305092

Thornton, C. A., & Smith, P. J. (1988). Action research: Strategies for learning subtraction facts. *Arithmetic Teacher*, *35*(8), 8–12.

Torbeyns, J., Verschaffel, L., & Ghesquiere, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction*, *23*, 1–21. http://dx.doi.org/10.1207/s1532690xci2301_1

University of Chicago School Mathematics Project. (2005). . *Everyday mathematics Teacher's Lesson Guide* (Vol. 1) Columbus, OH: McGraw-Hill.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte, NC: NCTM & Information Age Publishing.

Wilburn, D. B. (1949). Methods of self instruction for learning arithmetic. In *Supplementary Educational Monographs [No. 70]*. Chicago, IL: University of Chicago.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594