

FOSTERING FLUENCY WITH BASIC ADDITION AND SUBTRACTION FACTS USING
COMPUTER-AIDED INSTRUCTION

BY

MICHAEL D. EILAND

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Elementary Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Arthur J. Baroody, Chair
Professor Sarah Theule Lubienski
Professor Michelle Perry
Assistant Professor David J. Purpura, Purdue University

ABSTRACT

Achieving fluency with the basic addition and subtraction combinations is difficult for many students in grades 1 to 3. Two papers chronicle three experiments, which entailed evaluating the efficacy of software designed to promote fluency with add-with-8 or -9, near-doubles (e.g., $5 + 6$), and subtraction items via moderately guided learning of reasoning strategies. In Experiments 1 and 2, eligible students were randomly assigned to either a guided make-10 condition (e.g., the sum of $9 + 7$: $9 + [1 + 6] = [9 + 1] + 6 = 10 + 6 = 16$), or guided near-doubles condition (e.g., the sum of $3 + 4$: $3 + [3 + 1] = [3 + 3] + 1 = 6 + 1 = 7$). Experiment 3 also included a third condition, guided subtraction training (e.g., the difference $12 - 9$ can be thought of as: What number when added to 9 equals 12?). Each experiment involved pupils in Grade 1, 2, or 3. ANCOVAs revealed each of the guided interventions promoted learning of the trained strategies as evidenced by transfer to unpracticed items.

ACKNOWLEDGEMENTS

First, I wish to thank my Lord and Savior Jesus Christ, who made the opportunity to pursue this degree possible. This journey has strengthened my faith.

I wish to thank Professor Emeritus Arthur J. Baroody. His vision, passion, and ability to look beyond traditional educational stances, promoted my growth both as a researcher and as a teacher. Dr. Baroody has allowed me to work with hundreds of students, many of whom as I encounter years later, are eager to share with "Mr. Michael" the things they learned from their time spent as research participants and more importantly, how much they enjoy mathematics now.

Thank you to Professors Sarah Lubienski, Michelle Perry, and David Purpura for being excellent committee members, advocates for my research efforts, providing dissertation guidance, and for launching my career. Professor Lubienski and Dr. Michelle Crockett, my first graduate school advisor, served as supervising professors during my days as a teaching assistant and gave me the confidence to reach beyond the text.

A special thank you to Dr. Violet J. Harris, without whose vision and support, my graduate experience at the University of Illinois Urbana-Champaign would not have been possible. I have been fortunate to work alongside some extremely talented fellow graduate students and Academic Professionals. I appreciate your willingness to collaborate as well as the hard work and the dedication you've shown through the years. I also extend thanks to my former pupils as well as the research participants who have so graciously allowed me to serve you through the years. Your collective efforts continue to inspire me.

To my grandmothers, who always made time to encourage me while instilling the value of assisting others would in turn make me a more fulfilled individual. Granny Mattie taught me

that “no knowledge is wasted” while Grandma Elnora even in her late 90s had a thirst and passion to learn something new each day. To my siblings, William Jr. and Tracee, thank you for never making me feel like a "baby brother" despite a significant difference in age, always providing an opportunity to tag along when appropriate. I love you both dearly. To my five nephews and three godchildren, all of whom have memories of Uncle Mike almost always being enrolled in college; understand that you can accomplish anything you put your mind to and effort toward. Last but not least I wish to thank my parents, who met in college and have remained married for over 50 years, are still a source of inspiration. Thank you to my father, William Sr., who as a stationary engineer, worked in crawl spaces on arthritic knees lacking cartilage, in some of the toughest Chicago housing developments, often placing his life at risk, for always putting his needs last so that I could have the opportunity to receive an education. Thank you to my mother, Gladys, who showed me that a dream delayed does not have to be a dream deferred, by graduating college in her 40s while working full-time and meeting family responsibilities. It is upon all of your shoulders I stand.

TABLE OF CONTENTS

Chapter I. FOSTERING FLUENCY WITH BASIC ADDITION COMBINATIONS WITH GRADE 2 STUDENTS

ABSTRACT.....	1
Rationale for the Present Study.....	6
Methods.....	10
Results.....	24
Discussion.....	31
Conclusion	33
References.....	35

Chapter II. FOSTERING ARITHMETIC FLUENCY WITH GRADE 1 AND GRADE 3 STUDENTS

ABSTRACT.....	44
Rationale for the Present Study.....	48
Experiment 1 – Grade 1 Students	51
Methods.....	53
Results.....	64
Discussion.....	70
Experiment 2 – Grade 3 Students	73
Methods.....	77
Results.....	84
Discussion.....	91
Conclusion	96

References	99
Appendix A: Curriculum Details	109
Appendix B: Mental Arithmetic Test Screen Shots.....	111
Appendix C: Procedures for Determining the Use of a Response Bias.....	112
Appendix D: Mental Arithmetic Training Screen Shots.....	114
Appendix E: Summary of Mental Arithmetic Training and Practice Provided by Curricula Used in the Schools.....	117
Appendix F: Experiment 1 Captions.....	118
Appendix G: Experiment 1 Response Biases Identification by Type, Condition, and Testing Session	122
Appendix H: Experiment 2 Screen Captions	123
Appendix I: False Positive Identification	127

Chapter I.
FOSTERING FLUENCY WITH BASIC ADDITION COMBINATIONS
WITH GRADE 2 STUDENTS

ABSTRACT

Achieving fluency with add-with-8 or -9 and near-doubles combinations to 20 is difficult for numerous grade 2 students. A 4-month training experiment entailed evaluating the efficacy of software designed to promote fluency via moderately guided learning of reasoning strategies. Seventy-six eligible students were randomly assigned to one of two conditions: structured make-10 (e.g., the sum of $9 + 7$: $9 + [1 + 6] = [9 + 1] + 6 = 10 + 6 = 16$) or structured near-doubles (e.g., the sum of $3 + 4$: $3 + [3 + 1] = [3 + 3] + 1 = 6 + 1 = 7$) for 30-minute sessions twice a week for 9 weeks. Each training condition served as an active control for the other condition. An ANCOVA revealed that at the delayed posttest, no statistical performance differences between the make-10 group and the near-doubles group on practiced and unpracticed add-with-8 or -9 combinations. The near-doubles group outperformed the make-10 group on practiced and unpracticed near-doubles combinations. Analyses of decreases in inefficient strategy use and increases in slow but correct responses indicated that both types of guided training promoted learning more effectively than the regular classroom instruction received by the active control.

A longstanding goal of elementary mathematics instruction is successful retrieval of basic addition items to 18 or 20 and the related subtraction items. Policy documents (Council of Chief State School Officers or CCSSO, 2010; Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics or NCTM, 2000, 2006) define success in terms of automaticity and fluency. Automaticity is accurate, quick, effortless, and non-conscious recall of facts. With the growing emphasis on meaningful learning or memorization, success has sometimes been defined more broadly as fluency (Baroody, Bajwa, & Eiland, 2009; cf. National Mathematics Advisory Panel or NMAP, 2008). A broad meaning of fluency involves the ability to appropriately, and adaptively, apply knowledge as well as efficiently remember it. Appropriate use implies applying knowledge to only relevant cases [e.g., $3 + 3 + 1$ is a suitable solution strategy for $3 + 4$, but not for $3 + 5$]. Adaptive use implies flexibly applying, transferring, or adjusting knowledge to solve new or moderately new problems. Although there is disagreement as to how to best attain the goal of efficiency or fluency with the basic sums, there is agreement that all children need to accomplish this early education milestone (CCSSO, 2010; Jordan, Hanich, & Kaplan, 2003; Kilpatrick, et al., 2001; NCTM, 2000, 2006; NMAP, 2008).

A pervasive characteristic of students with difficulties learning mathematics (DLM) is lacking fluency with the basic number combinations (Ackerman, Anhalt, & Dykman, 1986; Geary, 1996; Goldman, Pellegrino, & Mertz, 1988; Jordan, Hanich, & Kaplan, 2003; Jordan, Hanich, & Uberti, 2003; Russell & Ginsburg, 1984). The prevailing instructional method in most schools focuses on achieving memorization of basic number facts by rote through repeated unstructured practice. Such instruction is based, at least implicitly, on Thorndike's (1922) tenet, the *law of frequency*. This law stipulated that the more two stimuli are presented together (e.g., the more frequently a child sees an arithmetic combination such as " $4 + 5$ " and the correct

answer “9”) the stronger the association between the two addends and the sum becomes—resulting in efficient recall of the correct answer when that particular number fact is presented. Outgrowth models such as the distribution-of-associations model from Siegler and Jenkins (1989) and its successors (e.g., Sharger & Siegler, 1998), for example, specify that in order to achieve efficient fact recall, thousands of instances with a number fact and its sum (e.g., $9 + 7 = 16$) are necessary to establish a consistent memory trace. Facts are assumed to accumulate associative strength independent of their related commuted partner (e.g., practice with $3 + 4$ was presumed to have no effect on $4 + 3$). The embodiment or such theories—teaching the basic facts via drill and extensive practice (e.g., classroom and homework worksheets familiar to several generations of students), often contributes to children’s difficulties with learning the basic combinations (e.g., Baroody, et al., 2009; Brownell, 1935). Although learning basic facts by rote may be effective in promoting the efficient recall of certain number facts, it is relatively ineffective in promoting fluency for several reasons:

1. it may lead to associative confusions;
2. the tedium of memorizing hundreds of basic combinations by rote is burdensome and even overwhelming for many children; and
3. the routine of unstructured practice does not lead to transfer on conceptually similar number combinations.

In brief, Fayol and Thevenot (2012) summarize the fundamental drawback of traditional instruction best “it could be argued that it is cognitively economic to store a reproductive process [reasoning strategy] rather than numerous individual associations (Baroody, 1983).”

Promoting fluency requires actively constructing knowledge bases through meaningful memorization or learning by recognizing patterns or using relations. Meaningfully memorizing

basic combinations entails linking conceptual, procedural, and factual knowledge (Hiebert & Carpenter, 1992; Hiebert & Lefevre, 1986; Katona, 1967; Mason & Spence, 1999; Moursund, 2002; Resnick & Ford, 1981; Rittle-Johnson, Siegler, & Alibali, 2001) in a “closely knit system of understandable ideas, principles, and processes (Brownell, 1935, p. 19).” This relational learning can promote a well-structured knowledge base (number sense) that along with structured purposeful practice encourages computational fluency with the basic number facts. Achieving such learning requires a paradigm shift in instructional organization so as to (a) promote the recognition of conceptual regularities (patterns or relations), (b) build upon students’ existing knowledge, and (c) instill the value of utilizing general principles and concepts. Meaningful memorization promotes retention and transfer, two areas students often struggle with, better than learning separated from organized conceptual regularities (Brownell, 1941; Brownell & Chazel, 1935; Gersten & Chard, 1999; Henry & Brown, 2008; James, 1958; Jordan, 2007; Katona, 1967; Kilpatrick et al., 2001; NMAP, 2008; Piaget, 1964; Skemp, 1978, 1979, 1987; Steinberg, 1985; Suydam & Weaver, 1975; Swenson, 1949; Thiele, 1938; Wertheimer, 1959).

Typically, the process of meaningfully learning the basic number facts progresses through three overlapping phases (Kilpatrick et al., 2001; Steinberg, 1985). In Phase 1—counting, children use objects or verbal counting to determine sums. In Phase 2—reasoning, children use known combinations and relations acquired in Phase 1 to consciously deduce the answer to an unknown sum. In Phase 3—retrieval, children can appropriately, adaptively, and efficiently produce sums from memory. Deliberate (conscious and somewhat slow) reasoning strategies (Phase 2) serve as a key bridge between using relatively inefficient counting strategies and efficient retrieval (Phase 3) in critical ways. One way is that they promote adaptive

expertise—well understood knowledge that can be applied appropriately, adaptively, and efficiently to familiar as well as to new tasks (Hatano, 2003); an organizing framework for learning and storing both practiced and unpracticed combinations (Canobi, Reeve, & Pattison, 1998; Dowker, 2009; Rathmell, 1978; Sarama & Clements, 2009). A second way is that reasoning strategies can in time become automatic thereby serving as the basis for fluent retrieval (Baroody, 1985, 1994; Fayol & Thevnot, 2012; Verschaffel, Greer, & De Corte, 2007). As noted by Dowker (1992), professional mathematicians rely on patterns and relations to deduce answers to problems posed for which they cannot immediately recall the correct answer; therefore, in certain situations reasoning strategies may be superior in efficiency to mental recall.

A comparison of the ways in which the traditional (associations) and the number sense (relational) viewpoints incorporate Phases 1 to 3 within instruction are summarized in Table 1. For a thorough overview of traditional passive storage views versus number sense active construction views regarding the basic number combinations see Baroody and Purpura (in press).

Table 1
Comparison of Traditional and Number Sense Viewpoints of Phases 1 to 3

	Traditional View	Number Sense View
Unit of learning	Association between an expression and its answer	Relational knowledge (patterns and relations)
Mechanism of learning	Memorization of individual facts by rote—Observing and practicing (strengthening) an association	Meaningful memorization—Discovering patterns and relations—e.g., connecting new information to existing knowledge
Mental representation of mental arithmetic expert	Efficient (fast + accurate) from memory	Efficient (fast + accurate) with appropriate adaptive thought
Retrieval process(es)	Single fact retrieval process	Multiple retrieval processes
Rate of learning	Achieved in short order	Achieved gradually
Role of Phases 1, 2, and 3	Phases 1 and 2 <i>impede</i> the development of Phase 3 and should be discouraged. Phase 3 accomplished solely via memory recall.	Phases 1 and 2 <i>inform</i> the development of Phase 3. Phase 2 (in time) can be as efficient as Phase 3.

Most computer-based applications geared toward learning basic number combinations involve drill and practice. Research on such programs indicates that some are more effective than traditional classroom instruction (Goldman & Pellegrino, 1986; Hasselbring, Goin, & Bransford, 1988; Koscinski & Gast, 1993; see review by Kulik & Kulik, 1991; but cf. Fuson & Brinko, 1985; Hativa, 1988; Vacc, 1992). Effective early intervention, in the form of computer programs which focus on patterns and relations, may help students with mathematical difficulties master the basic addition combinations (Dev et al., 2002; Sarama & Clements, 2009).

Rationale for the Present Study

The purpose of the present training study was to evaluate the efficacy of experimental programs designed to foster a make-10 reasoning strategy or a near-doubles reasoning strategy. Research suggests that larger sums are more challenging to solve than smaller ones (Kraner, 1980; Smith, 1921; see Cowan, 2003, and NMAP, 2008, for reviews). A previous training experiment by Baroody, Thompson, and Eiland (2008) with advanced grade 1 students targeting make-10 and near-doubles reasoning strategies failed to foster fluency with larger sums. Discussed in turn are the reasons for the poor performance, upgraded features of the present intervention, research aims, and hypotheses.

Previous Training Efforts

Item Presentation. In the first make-10/near-doubles intervention (Baroody, Thompson, & Eiland, 2008), a brick (*Castle Wall Game*) or boxcar (*Train Game*) appeared in the upper center portion of the screen. For the make-10 condition, during subset A, students were presented with an item with the sum of ten, such as $9 + 1$, a three addend item $9 + 1 + 4$, the corresponding add-with-8 or -9 item $9 + 5$, its commuted partner $5 + 9$, and an unrelated filler item $5 + 7$. The program then presented a new targeted add-with-8 or -9 item beginning with the $10 + n$ item that

eventually be used to solve a plus-8 or plus-9 item while using the make-10 reasoning strategy. The remaining four problems follow the structure previously outlined. In the near-doubles condition, students were presented with a double, such as $3 + 3 = ?$, a three-addend item $3 + 1 + 3 = ?$, the corresponding near-double $4 + 3$, its commuted partner $3 + 4$, and the filler $3 + 5 = ?$; the remaining five problems within the subset have the same structure. Each condition completed two sets of 10 items within a session. Training was done using the same structure as the present study with the same computer feedback games (*Castle Wall Game* and *Train Game*). Training consisted of 16 sessions and lasted approximately 8 weeks. On incorrect responses, the child was encouraged by project staff to determine the correct sum using a physical manipulative. Each participant was exposed to the correct answer to their intervention specific trained items the same number of times. Preliminary ANCOVA results indicated no significant differences in fluency rate performance over standard classroom practice on add-with-8 or -9 or near-doubles items.

Some of the reasons for the lackluster performance by both groups include:

1. The negligibly guided instruction ordered targeted add-with-8 or -9 or near-doubles items in a particular sequence. Although addition items and commuted partners appeared consecutively, the targeted reasoning strategy was not explained during the intervention.
2. Unless the combination was answered incorrectly, there was no student participation or input beyond answering the presented item.
3. No feedback concerning why these particular items were presented together or prompts offering ways in which students could build connections among them.

Features of the Current Experimental Training

General Program Improvements. Refinements were made to address the lack of student involvement prevalent in the previous program (Baroody, Thompson, & Eiland, 2008):

1. Introduction of Stages with a specified goal varying the level of student engagement in forming the decomposition intermediary steps inherent within their intervention. Items are presented both symbolically and concretely using 10-frames.
2. A new Paint Game required students to form all addition pairs for two targeted addends that will be decomposed within their intervention and used to assist in the forming of either a ten or a double.

Measuring Competency. Previous research efforts (Baroody, Eiland, Thompson, 2009; Baroody, Eiland, Purpura, & Reid, 2012, 2013; Baroody, Purpura, Eiland, & Reid, 2014) used fluency rate (correct, fast, non-counted responses) as the sole measure of performance. While the importance of achieving fluency cannot be overstated, using the fluency rate as the only measure of achievement masked numerous students' growth progressions from the pretest to the delayed posttest. Teachers, trainers, and administrators anecdotally commented on the growth they observed in individual students but until now there was no measure available to quantify overall performance.

A 6-point fluency index (F-Index) was developed to clarify where a student's performance lies along the three phases of mathematical development (counting, reasoning, and fluency). In essence, the F-Index provides a complimentary overall performance measure by describing all correct responses based on strategy and response time. Like the fluency rate, false positives due to a response bias are treated as incorrect responses.

Aims of the Research and the Hypotheses Tested

A semester long training experiment with grade 2 students served to evaluate the efficacy of two experimental programs for fostering reasoning strategies. (The Institute of Educational Sciences, which sponsored this study, outlines the definition and requirements for efficacy at http://ies.ed.gov/funding/pdf/2012_84305A.pdf.) Participants were randomly assigned to a training condition involving either a make-10 or a near-doubles reasoning strategy. The make-10 training condition served as an active control for the near-doubles condition (controlled for, e.g., regular classroom training with near-doubles combinations), and the near-doubles training condition served as an active control for the make-10 condition (controlled for, e.g., regular classroom training with add with-8 or -9 combinations). The meaningful learning and practice of a reasoning strategy over a period of 9 weeks should impact mental-addition performance in the following three ways:

1. Some children may achieve automaticity or fluency (i.e., achieve Phase 3) with the strategy and thus improve their retention of practiced targeted items and transfer to conceptually similar unpracticed items (Buckingham, 1927; Olander, 1931).
2. Many second graders may learn the reasoning strategy but not achieve fluency because of the limited amount of practice (i.e., take the important step of achieving Phase 2). These children should be able to apply the strategy appropriately and adaptively, if only deliberately (slowly), to practiced and, more importantly, to related but unpracticed items.
3. As a result of the targeted group's greater use of automatic retrieval (Phase 3) and deliberate application of the practiced strategy (Phase 2), there should be a significant

drop in the use of counting (Phase 1) strategies, very slow (unidentified) strategies, and incorrect responses on practiced and unpracticed targeted combinations (but not items emphasized in the other intervention).

Hypotheses

The following three hypotheses were evaluated:

Hypothesis 1: The make-10 training should facilitate the meaningful learning of a general make-10 addition reasoning strategy above and beyond regular classroom instruction received by its active control (the near-doubles group). At the delayed posttest, the make-10 group should have significantly better mean fluency index (F-Index) and mean fluency rate performance for the practiced and unpracticed add-with-8 or -9 items than the active control group.

Hypothesis 2: The near-doubles training should facilitate the meaningful learning of a general near-doubles addition reasoning strategy above and beyond regular classroom instruction as represented by the make-10 group. At the delayed posttest, the near-doubles group should have significantly better mean fluency index (F-Index) and mean fluency rate performance for the practiced and unpracticed $n + (n + 1)/(1 + n) + n$ items than the make-10 group.

Hypothesis 3: Success with the doubles items should lend itself to success with the near-doubles items. At the delayed posttest, is success with the doubles items, as determined by the fluency rate measure, a necessary condition for success on the near-doubles items?

Methods

Participants

Participants were recruited from 14 grade 2 classrooms in four elementary schools from adjacent school districts serving two, medium-sized, mid-western cities during Fall, 2007.

Parental consent forms were returned for 92 students. Fourteen students did not participate in the

study because they either moved prior to participant assignments ($n = 1$), scored too low on mental-arithmetic testing and subsequently assigned to a study emphasizing addition-with-1 and small near-doubles items ($n = 11$), did not complete training ($n = 1$), or joined the study late and had limited proficiency with the adding-with-1 and near-doubles items ($n = 1$). Study participants met the following criterion, fluent (correct by an undetermined strategy or a reasoning strategy in < 3 seconds) on 25% to 50% of the add-with-8 or -9 items on the pretest. Two students were too advanced and tested out of the study. Among the 76 participating students who completed the study (7.0 to 9.7 years of age; mean = 7.6 years; median = 7.5 years old; SD = 0.5 years), 55.3% were male. The composition of the sample was 36.8% African-American; 35.5% Caucasian; 13.2% Multi-ethnic, 9.2% Hispanic; and 5.3% Asian, respectively. Additionally, 61.8% of the participants were eligible for free or reduced lunch. The participant demographic included students who were teacher-identified and other marginalized populations. They were chosen because of the high likelihood they would not be fluent with the basic sums and might benefit from the intervention. Risk factors were defined as: eligible for free/reduced lunch, English as Second Language, or documented medical condition. See Table 2 for the demographic information by condition.

Table 2
Study 1 Participant Characteristics by Condition

		Training Condition	
		Structured Make-10	Structured Near-doubles
Age range		7.1 to 9.7	7.0 to 8.7
Mean (SD)		7.6 (0.5)	7.6 (0.4)
Median age		7.5	7.5
School			
	1	6	7
	2	8	7
	3	14	12
	4	11	11
Number of boys/girls		21 / 18	21 / 16
Free/Reduced lunch eligible		26	21
Black/Hispanic/Multiracial		24	21
English as Second Language		2	8
Medical Condition			
	ADHD	1	1
	Birth Complications	1	1
	Fetal Alcohol Syndrome	1	0
	Low Birth Weight	1	0

All four participating schools were committed to achieving the State's grade 2 learning objectives that included operations on whole numbers such as solving one- and two-step problems and performing computational procedures using addition and subtraction <<http://www.isbe.net/ils.math/capd.htm>>. The 10 classrooms in Schools 1, 2, and 3 used *Everyday Mathematics* (University of Chicago School Mathematics Project or UCSMP, 2005). The four classrooms in School 4 used *Math Expressions* (Fuson, 2006) and supplemented the teaching of computation with the *Touch Math* curriculum. Details regarding how these curricula approached mental arithmetic, in general, and add-with-8 or -9 items, in particular, can be found in Appendix A. All 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 structured discovery software for young children, it is likely these programs focused on drill of number skills.

Project hired personnel consisted of four female and one male Academic Professionals (APs) and three female and three male Research Assistants (RAs). Five APs and two RAs had previous teaching experience (1 to 17 years with a median = 5.5 years and 2 to 30 years with a median = 2 years, respectively). All had a B.S. or B.A.; seven majored in education and two in psychology or counseling. Four APs and all four RAs had a M.Ed. Four APs and one RA had teaching certificates. Four of the APs and two of the RAs had previous experience on the project. Prior to the beginning of the study all staff members had six 3-hour training sessions on testing/training procedures.

Administrative Procedure. All testing and training was done on a one-to-one basis with a project staff member typically in 30-minute sessions, twice a week. Testing and training were conducted at project-dedicated workstations (Apple G4 desktops or MacBook laptops) in a designated research space (e.g., shared reading room or auxiliary classroom) or a hallway outside the students' classroom. Unlike previous or follow-up studies, there was no preparatory training common to all students (Baroody, Eiland, & Thompson, 2009; Baroody, Eiland, Purpura, & Reid, 2013, 2014). During the first 2 weeks of the project, students were administered the pretest in order to establish baseline performance with adding-with-8 or -9, near-doubles, and filler items. Participants were randomly assigned to either the make-10 intervention or the near-doubles intervention with groups balanced by performance on all add-with-8 or -9 items. A delayed posttest featuring sets 1 to 4 identical to the pretest and reordered pretest items for sets 5 to 8 was administered 4 weeks after the conclusion of the training.

Measures

Mental-arithmetic testing. The test of addition fluency included seven categories of items as outlined in Table 3:

Table 3
Study 1 Tested Items by Combination Category

Combination Family	Combinations	Practiced in Stages I to IV by Group	
		Structured Make-10	Structured Near-doubles
Practiced add-with-8 or -9 items	$5 + 8, 5 + 9, 7 + 9, 8 + 5, 9 + 5, 9 + 7$	Yes	No
Transfer add-with-8 or -9 items	$4 + 8, 6 + 8, 6 + 9, 8 + 4, 8 + 6, 9 + 6$	No	No
Practiced doubles items	$3 + 3, 4 + 4, 5 + 5, 6 + 6, 7 + 7, 8 + 8$	4 + 4 and 8 + 8 Yes	Yes
Practiced Near-doubles items	$3 + 4, 4 + 3, 5 + 6, 6 + 5, 6 + 7, 7 + 6$	No	Yes
Transfer doubles and near-doubles items	$2 + 2, 2 + 3, 3 + 2, 4 + 5, 5 + 4$	No	No
Practiced filler items	$2 + 7, 3 + 5$	Yes	Yes
Transfer filler items	$1 + 1, 1 + 2, 2 + 1, 3 + 1$	No	No

Note 1. Prior to the mental-arithmetic test, students were given a pretest in order to introduce them to task expectations.

Set 0 in order was: $9 + 0, 4 + 1, 6 + 3, 5 + 2, 9 + 1, 5 + 7, 0 + 4, 1 + 7, 6 + 0$, and $1 + 3$.

Set 00 in order was: $8 + 1, 7 + 4, 3 + 0, 5 + 1, 1 + 9, 2 + 4, 0 + 8, 1 + 6, 4 + 6$, and $0 + 5$.

Note 2. The mental-arithmetic pretest (Sets 1 to 4)/posttest (Sets 1 to 8) was composed of sets of 10 items each.

Set 1 in order was: $5 + 5, 2 + 3, 9 + 6, 5 + 4, 7 + 6, 3 + 4, 8 + 5, 1 + 1, 4 + 8$, and $7 + 9$.

Set 2 in order was: $5 + 6, 8 + 6, 3 + 5, 1 + 2, 8 + 9, 6 + 6, 5 + 9, 4 + 4, 3 + 1$, and $8 + 7$.

Set 3 in order was: $2 + 2, 6 + 8, 2 + 7, 9 + 5, 3 + 3, 9 + 8, 6 + 5, 9 + 9, 7 + 8$, and $4 + 5$.

Set 4 in order was: $3 + 2, 7 + 7, 5 + 8, 4 + 3, 6 + 9, 8 + 4, 9 + 7, 2 + 1, 8 + 8$, and $6 + 7$.

Set 5 in order was: $2 + 7, 3 + 5, 8 + 7, 5 + 6, 9 + 9, 2 + 2, 5 + 9, 8 + 9, 1 + 2$, and $8 + 6$.

Set 6 in order was: $4 + 8, 7 + 9, 5 + 4, 8 + 8, 2 + 3, 7 + 7, 8 + 5, 9 + 6, 3 + 4$, and $7 + 6$.

Set 7 in order was: $6 + 9, 4 + 4, 6 + 7, 4 + 3, 1 + 1, 5 + 8, 3 + 1, 8 + 4, 3 + 2$, and $9 + 7$.

Set 8 in order was: $6 + 5, 7 + 8, 3 + 3, 9 + 8, 2 + 1, 9 + 5, 6 + 6, 5 + 5, 6 + 8$, and $4 + 5$.

Mental-arithmetic testing was placed in the context of a computer game. Illustrations, details, and instructions on the computer-assisted testing procedure are provided in Appendix B in the context of the *Car Race Game*. The tested items were presented in a partially random order with the following constraints: (a) two items of the same type (e.g., near-doubles), (b) containing the same addend, or (c) the same sum did not appear on consecutive trials. Prior to beginning the actual pre/posttest a practice test featuring untested items of interest acclimated students to task

expectations. The test consisted of four sets of 10 items each. Each pre/post testing session consisted of a testing set, a reward game, a second testing set, and a final reward game.

Mental-Arithmetic Scoring. Scoring involved a two-step process: (a) scoring each trial on the mental-arithmetic test and (b) using the relevant trial scores to determine a mean composite score for a family of combinations.

Scoring trials. Scoring of a trial took into account three factors: (a) accuracy, (b) reaction time, and (c) strategy.

Accuracy. Correct answers as well as spontaneous corrections. For example, if a child responded to $7 + 8$ with, "Fourteen, [immediately] no wait, fifteen," the child was credited with a correct response. However, when a child gave the correct sum as part of a range of estimates (e.g., for $7 + 8$, saying "It could be thirteen, fourteen, or fifteen") or a series of wild guesses (e.g., for $8 + 6$, saying "Twelve?" "Fifteen?" how about "Fourteen?"), the response was scored as incorrect. Irrelevant or non-numerical answers were also scored as incorrect. Finally, false positives due to a response bias were scored as incorrect.

Some mental-addition novices state a consistent pattern of responses without regard to the addends involved (Baroody, 1999; Dowker, 2003). For example, some children simply state a particular number repeatedly (e.g., $5 + 8 = 15$, $6 + 7 = 15$, $7 + 7 = 15$, $6 + 9 = 15$), or misuse a reasoning strategy (e.g., $5 + 8 = 18$, $6 + 7 = 17$, $10 + 8 = 18$, $6 + 9 = 19$) and accidentally stumble upon a correct sum. In order to guard against the non-discriminate application of an inappropriate strategy, trials were scored in the context of a child's answer to other items during a testing session. Specifically, a response bias determination was done using the total number of items in a 2-set testing session. A complete description of the process of determining a response bias complete with an example can be found in Appendix C.

Response time. Definitions for automaticity, fluency, and proficiency all include an element concerning the time it takes to respond to a presented combination. There remains no single standard for a “fast” response time. For example, Torbeyns et al. (2004) used a 2 second response time interval for identifying fast answers while Siegler (1988) allowed participants up to 5 seconds. Although Threlfall, Frobisher, and MacNamara (1995) noted that “[r]esponse time is not a clear and unambiguous indicator of the strategy used to arrive at the answer to addition questions.” Bajwa-Priya (2013) investigated the impact of response time criteria on the number of combinations that would be deemed retrieval (fast and accurate). She found that a 3-second criterion separated overt counting or reasoning (Phase 1 or 2) strategies from covert strategies (indicative of Phase 3 strategies) for both addition and subtraction trials. Therefore, for the purposes of this study, *fast* responses are operationally defined as those generated in less than 3 seconds. In instances where either overt reasoning or counting was used to generate a response in less than 3 seconds, the answer was coded according to strategy. Responses generated between 3 to less than 6 seconds are defined as *deliberate*, between 6 to less than 15 seconds are defined as *slow*, lastly 15 or more seconds are defined as *no response*. The less than 3-second interval for fast responses was found to be an effective delineation for both addition and subtraction items in previous research (Baroody, 1999; Dowker, 2003; Priya, 2013).

Strategies. The four distinct strategies identified and outlined in Table 4 were:

Table 4
Response Categories for Mental-Arithmetic Test

Category	Response Type
Reasoning	Overt evidence of logically deducing a response using known facts or relations (for $8 + 6$, e.g., stating: “Eight plus two is ten. Ten plus four is fourteen.”).
Counting	(a) Counting objects concretely; (b) verbally stating a portion of the count sequence; or (c) by sub-vocally using a mental representation of the count sequence accompanied by movement of the head, the fingers, or the eyes.
Non-determined	No evidence of counting, reasoning, or other overt figuring
No Response	No answer before 15 seconds elapsed.

Composite scores (dependent measures). Described in turn are the two dependent measures used in the analyses, F-Index and fluency rate.

F-Index. The fluency index (F-Index) was designed to gauge overall progress towards efficient strategy use, i.e., fluency. This 6-point scale parallels the three phases of mental-addition development previously discussed. The F-Index rationale is outlined in Table 5:

Table 5
F-Index Developmental Phase Correspondence and Response Characteristics

Points	Developmental Phase	Response Characteristics
5	Phase 3	Correct, automatic (< 3 seconds) fact retrieval (fact recall or unconscious reasoning strategy)
4	Transition from Phase 2 to Phase 3	Correct, fast (< 3 seconds) overt reasoning
3	Phase 2	Correct, deliberate ($3 \leq x < 6$ seconds) overt reasoning
2	Transition from Phase 1 to Phase 2	Correct, slow ($6 \leq x < 15$ seconds) reasoning or deliberate undetermined (other) strategy
1	Phase 1	Correct, slow undetermined strategy or counting
0	Pre-Phase 1	False positive due to a response bias, no response, or incorrect response

A child’s F-Index score for a combination family was the mean of the 0 to 5 ratings assigned to each tested practiced or transfer item.

Fluency Rate. Fluency rate is based on a two-point scale: 1 = a fluent (correct, fast, and non-counted) response and 0 = non-fluent response (correct sum in 3 or more seconds or an

incorrect sum). The fluency rate is the mean of fluent responses (i.e., responses scored as either 4 or 5 using the F-Index scale) for a combination family.

A macro developed by the author checked all items for correctness, assigned scores for F-Index and fluency rate, adjusted scores for response biases, and produced means by combination family.

Training Intervention

If the teaching of reasoning strategies via a computer program is viewed along a continuum from *unguided* instruction (no scaffolding or connections) to *highly guided* instruction (explicitly illustrating and explaining how the strategy works, depicting how the strategy is useful for solving targeted items, and providing interactive opportunities for the student to engage all facets of learning the strategy), then the structure of Study 1 falls somewhere in the middle of these two extremes (see Baroody, Purpura, Eiland, & Reid, in preparation; for a more thorough explanation). The curricular emphasis for Study 1 is best described as *moderately guided* instruction. In relation to make-10 and near-doubles trainings, *moderately guided* instruction features: (a) the commuted item immediately following the decomposition illustration of the targeted strategy; (b) the child actively performs the intermediary steps within a strategy by clicking on the number of dots necessary to compose a ten or create a doubles; and (c) the participants progress through a sequence of pre-planned lessons that involve making-a-ten or decomposing the larger addend of a near-doubles into a doubles plus one item. Unlike *highly guided* instruction, in *moderately guided* instruction, there is no explicit specification of a target strategy, no explicit prompt asking if a particular relation helps to solve a targeted strategy, and most importantly, no explanation of why the reasoning

strategy is useful for particular arithmetic combinations. Discussed in turn, are the preparatory training, the role of the project trainers, and the fidelity of the intervention.

Preparatory Training. Training consisted of 20 sets divided into four Stages and lasted 9 weeks. Each set was divided into three subsets, subset A—the *Train Game*, subset B—the *Castle Wall Game*, and subset C—the *Paint Game*. In Stage I (Sets 1 to 4), a targeted make-10 (near-doubles) item is displayed symbolically. The tester reads the problem (e.g., $9 + 5$) “How much is $9 + 5$?” After the student responds, the tester stops the clock with the **ENTER** key on the number pad and types in the response. A fully automated process engaged displaying how the make-10 (near-doubles) strategy can be used to determine a sum. The addends are concretely represented above the numerals in color-coded five/ten frames, the first addend in blue, the second addend in red. Following the entering of the response, 1 dot (yellow) from the smaller addend (5), jumps between the two five/ten frames, leaving 4 as a decomposed addend. The revised problem $9 + 1 + 4$ is displayed. Next, the computer displayed $9 + 1 = 10$, followed by $10 + 4$, and finally, the equation $10 + 4 = 14$. For incorrect responses, no feedback is given regarding the response. During Stage I, commuted items are presented consecutively. After all of the items within the first subset (Set A) have been completed, a train engine storage yard appears (*Train Game*). Items with identical sums appear on the same vertical train track. Students’ accuracy with the targeted items is displayed in terms of the number of stars gained out of 10. The feedback for the second subset (Set B) of Stage I was similar; equivalent sums were displayed within the same sections of a horizontally arranged castle wall (*Castle Wall Game*). The *Paint Game* was the third and final subset (Set C) within a daily session. To complete the game, a student must find all blue/red combinations that sum to the targeted number. Students are presented with a target number (5 or 7 for make-10 training; 4 or 8 for near-doubles training) that will be decomposed

during Set A and Set B of their intervention such that the targeted addend will assist in the forming of ten or a doubles. A button representing the blue paint is pressed and a student fills in circles to depict the first addend. Similarly, a button representing the red paint is pressed and the remaining circles on a row are filled in to depict the second addend.

In Stage II (Sets 5 to 12), selected add-with-8 or -9 and near-doubles items prompt students to respond to the decomposition of the smaller addend (make-10) or larger addend (near-doubles) component of the reasoning strategy. The goal was to have students actively attend to the steps necessary to successfully use the reasoning strategy within their intervention. Specifically, it provided the opportunity to discuss the elements required to form the needed ten or doubles in order to successfully use the reasoning strategy within their trained intervention. Non-designated problems required a single initial response prior to the computer automatically displaying the targeted reasoning strategy as in Stage I. Using $3 + 4$ as an example: the tester read the problem “How much is $3 + 4$?” After the student responded and the tester enters the information, the computer displayed $3 + 4$ symbolically and represents the 3 using blue dots and 4 using red dots in a separate color-coded five frames. The tester prompts “Click on the number of dots needed to turn 4 (larger addend) into 3.” Using the mouse, the student indicates how many dots should be removed from the 4 in order to make 4 into 3; subsequently 1 yellow dot jumps from the frame containing 4 and lands between the frames. The tester asks “ $4 - 1$ equals?” Following the response, the computer displayed the difference in blue dots. The 1 dot taken from the addend 4 appears in yellow, jumps in front of the first addend (3), and $1 + 3 + 3$ is displayed. The student is asked to solve the doubles portion “ $3 + 3$ ”; after responding, the sum (6) is displayed in orange dots. Finally, after solving $6 + 1$, the 6 orange dots and the 1 yellow dot are

displayed in a single ten-frame. Like Stage I, commuted items are presented consecutively.

Table 6 details the designated problems and the decomposition transitions for each strategy.

Table 6:
Study 1 Decomposition of Selected Targeted Items in Stages II and III by Condition

Training	Targeted Item	Student Answered Decomposition Transitions		
		1	2	3
Add-with-8 or -9	5 + 8	? + 8 = 10	5 - 2 = ?	3 + 10
	8 + 5	8 + ? = 10	5 - 2 = ?	10 + 3
	5 + 9	? + 9 = 10	5 - 1 = ?	4 + 10
	9 + 5	9 + ? = 10	5 - 1 = ?	10 + 4
	7 + 8	? + 8 = 10	7 - 2 = ?	5 + 10
	8 + 7	8 + ? = 10	7 - 2 = ?	10 + 5
	7 + 9	? + 9 = 10	7 - 1 = ?	6 + 10
	9 + 7	9 + ? = 10	7 - 1 = ?	10 + 6
Near-doubles	3 + 4	4 - ? = D	3 + 3	6 + 1
	4 + 3	4 - ? = D	3 + 3	1 + 6
	5 + 6	6 - ? = D	5 + 5	10 + 1
	6 + 5	6 - ? = D	5 + 5	1 + 10
	6 + 7	7 - ? = D	6 + 6	12 + 1
	7 + 6	7 - ? = D	6 + 6	1 + 12
	7 + 8	8 - ? = D	7 + 7	14 + 1
	8 + 7	8 - ? = D	7 + 7	1 + 14

In Stage III (Sets 13 to 16), students no longer are required to click on the number of dots necessary to form ten or form a doubles. Like Stage II, the tester asks prompts concerning the number of dots necessary to make-10 or form a doubles. No corrective opportunities are provided when an initial response is incorrect. In Stage IV (Sets 17 to 20), students no longer decompose an addend. After the initial response, the computer displays the make-10 or near-doubles strategy without explanation in a fully automated manner like Stage I. Unlike Stage I, commuted items are not presented back to back. See Appendix D for screenshots of the *Train Game*, the *Castle Wall Game* and the *Paint Game*.

Following a Shortcut task that gauged whether or not students referenced a previous problem to answer a presented item and Winter Break, students were administered the mental-arithmetic delayed posttest. Sets 1 to 4 of the mental-arithmetic posttest were identical to the

pretest; Sets 5 to 8 reorganized the pretest items while maintaining the rules of no consecutive items having the same sum, the same addend, or belongs to the same combination family. A typical training/testing day consisted of a subset followed by a student chosen reward game and a final subset(s) of material.

Role of the Project Trainers. Administrative duties included completing lesson and mental-arithmetic test log sheets to ensure each participant completed each lesson or testing session without duplication. Logistical duties included picking up children from their classrooms, obtaining positive assent for each testing and training session, escorting children to a project workstation, logging them into the appropriate training program, and encouraging on-task behavior. The trainer's primary instructional role was to give voice to the scripted instructions, read number sentences during mental-arithmetic testing, and explain feedback graphically displayed by the computer screen.

Fidelity of Intervention. Fidelity of training was ensured by (a) 10 hours of staff training prior to the beginning of the project on the rationale for the programs and procedures for implementing the computer guided training, (b) a copy of the Trainer Guidelines at each computer station, (c) brief (10 to 30 minute) staff meetings during the semester to review procedures and address training issues as needed, and (d) a lesson log sheet for keeping track of which lessons each student had completed. The programs ensured that each child received (a) the assigned intervention (upon logging in the child was automatically connected to his or her treatment), (b) the combinations in the order specified by an intervention, and (c) feedback on correctness. In regard to implementation fidelity, all students saw the correct answer the same number of times and completed all 20 sets before the posttest. Each targeted item was practiced 20 times during the 9 weeks of training.

The make-10 and near-doubles groups practiced mutually exclusive items therefore, served as an active control for each other. Various threats to internal validity were accounted for by the random assignment. Significant posttest differences cannot be attributed to history (e.g., classroom instruction or practice), regression to the mean, maturation, or selection, because theoretically random assignment ensures all groups are comparable on these confounding variables. Students received identical tests the same number of times regardless of training condition to discount a testing effect. Both groups received identical *Castle Wall Game*, *Train Game*, and *Paint Game* training interfaces as well as identical reward games via the computer to control for a novelty effect. Any contamination or diffusions effect, which would facilitate the learning of the near-doubles and add-with-8 or -9 items by sensitizing the child to the mathematical regularities adds to measurement error and makes it more difficult to obtain significant results.

Analytic Procedure

ANCOVAs, using experimental condition as the grouping variable and mental-arithmetic pretest scores, age, and free/reduced lunch eligibility as the covariates, were used to compare mean proportion correct posttest performance of each group on targeted practiced and transfer combinations. The main intervention group was compared to the active control group for each of the two primary analysis sets (add-with-8 or -9 and near-doubles). A correction (Benjamini & Hochberg, 1995) was applied to adjust for Type I error due to multiple comparisons. The adjustment was applied separately for the near-doubles and add-with-8 or -9 items. For each, there were a total of two comparisons (practiced items and transfer items). Additionally, effect size magnitude (Hedges' g) was examined for all contrasts due to the limited power of the study and the importance of evaluating effect sizes (Wilkinson & APA Task Force on Statistical

Inference, 1999). Effect sizes were calculated after accounting for the covariates using the posttest mean proportion correct. As the significance level does not necessarily lead to a relationship of “practical significance or even to the statistical magnitude of the effect” (Lipsey et al., 2012, p. 3), per the Institute of Education Science’s (IES) What Works Clearinghouse (WWC) “effects that are not statistically significant but have an effect size of at least 0.25 are considered ‘substantively important’” (IES, 2014, p. 23). A Hedges’ g that exceeds 0.25 will serve as the effect size measure to evaluate whether a particular intervention is substantively important and efficacious. In cases where Levene’s Test of Equality of Error Variances was violated, a condition*pretest interaction was added into the model. Gender, school, and classroom were initially included in the analysis as random variables. Since none of these variables were statistically significant, they were removed from all analyses. Thus all reported analyses feature experimental condition as the only random variable.

Two students on the posttest consistently stated a favorite number during Session 3. The mean combination performance for these students was adjusted to account for the false positives (4 items total). A complete summary of false positive performance can be found in Appendix C.

A significant p value on McNemar’s test was used to confirm whether or not prerequisite knowledge is necessary for success on targeted items. McNemar analyses were conducted within groups for each prerequisite / targeted item pair.

Results

Descriptive statistics for children’s performance on each of the targeted combination categories are presented in Table 7 and Table 8. Pretest results, and the delayed posttest results regarding Hypotheses 1 to 4 are discussed in turn. All reported statistically significant posttest results for main effects of condition remained so when the Benjamini-Hochberg adjustment for

multiple comparisons was applied. Given the directional nature of the contrasts, effects of treatment were tested using 1-tailed significance values unless noted.

Pretest Analyses

Preliminary analyses revealed that the composition of children in the two groups did not differ in gender, $\chi^2(1, N=76) = 0.65, p = .799$; free/reduced lunch eligibility, $\chi^2(1, N=76) = 0.79, p = .374$, or ethnicity (comparing African American, Asian, Caucasian, Hispanic, and Multi-racial), $\chi^2(4, N=76) = 2.56, p = .634$. Using proportion correct, 2-tailed ANOVAs revealed no significant differences among groups for the targeted all add-with-8 or -9 items, $F(1, 74) = 0.18, p = .673$, and all near-doubles items, $F(1, 74) = 1.27, p = .263$, respectively. Cronbach's alpha for all add-with-8 or -9 items and all near-doubles items were $\alpha = .50$ and $\alpha = .72$, respectively.

Hypothesis 1: Efficacy of the Structured Make-10 Training

F-Index Measure. For add-with-8 or -9 items (practiced by the make-10 group), after applying the Benjamini-Hochberg adjustment, there were no significant differences between groups on practiced items, $F(1, 70) = 1.37, p = .123$, Hedges' $g = 0.64$, or transfer items, $F(1, 71) = 0.33, p = .284$, Hedges' $g = 0.15$. Conceptually similar items $7 + 8/8 + 7$ practiced by both groups and $8+9/9+8/9+9$ unpracticed by both groups, produced results of, $F(1, 70) = 1.75, p = .190$ (2-tailed), Hedges' $g = -0.38$ and $F(1, 67) = 0.38, p = .538$ (2-tailed), Hedges' $g = -0.06$, respectively. In the case of both types of practiced items, the effect sizes exceed the IES (2014) Hedges' $g = 0.25$ criteria for substantively important practice favoring the make-10 intervention on add-with-8 or -9 items, but favoring the near-doubles intervention on $7 + 8/8 + 7$.

Fluency Rate Measure. On items scored as either 4- or 5-points on the F-Index, after applying the Benjamini-Hochberg adjustment, the structured make-10 group significantly outperformed the near-doubles group on practiced add-with-8 or -9 items, $F(1, 70) = 7.57, p$

=.004, Hedges' $g = 0.67$ but not on the transfer items $F(1, 71) = 0.41, p = .261$, Hedges' $g = 0.14$. For the practiced $7 + 8/8 + 7$ items, after applying the Benjamini-Hochberg adjustment, the near-doubles group marginally outperformed the make-10 group, $F(1, 71) = 4.83, p = .031$ (2-tailed), Hedges' $g = -0.51$, with no statistical differences on unpracticed $8 + 9/9 + 8/9 + 9$ items, $F(1, 67) = 0.12, p = .732$ (2-tailed), Hedges' $g = -0.07$. On practiced add-with-8 or -9 items, the effect sizes exceed the IES (2014) Hedges' $g = 0.25$ criteria for substantively important practice favoring the make-10 intervention, but favoring the near-doubles intervention on $7 + 8/8 + 7$.

Hypothesis 2: Efficacy of the Structured Near-Doubles Training

F-Index Measure. For near-doubles items (practiced by the near-doubles group), after applying the Benjamini-Hochberg adjustment, the near-doubles group significantly outperformed the make-10 group on practiced items, $F(1, 70) = 8.20, p = .003$, Hedges' $g = 0.72$, and more importantly, on transfer items, $F(1, 71) = 8.59, p = .002$, Hedges' $g = 0.45$.

Fluency Rate Measure. After applying the Benjamini-Hochberg adjustment, the results were consistent with those of the F-Index measure. The near-doubles group significantly outperformed the make-10 group on practiced items, $F(1, 70) = 7.59, p = .004$, Hedges' $g = 0.77$, and transfer items, $F(1, 71) = 11.59, p < .001$, Hedges' $g = 0.52$. The effect sizes for both practiced and transfer items on the F-Index and the fluency rate measures exceed the IES (2014) Hedges' $g = 0.25$ criteria for substantively important practice favoring the near-doubles intervention.

Hypothesis 3: Developmental Prerequisite Knowledge

Regardless of intervention, on all fourteen comparisons, students who were not fluent on prerequisite doubles item were not fluent on the corresponding near-doubles item for both

practiced and transfer items on the posttest. Using McNemar's test, all 1-tailed comparisons indicate p-values $< .05$. See Table 9 for a summary of performance.

Table 7
Study 1 Pretest and Adjusted Posttest F-Index and Fluency Rate Scores by Condition

Condition	F-Index (0 to 5)				Fluency rate (0 to 1)			
	Pretest		Adjusted ^a Posttest		Pretest		Adjusted ^a Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Practiced Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Practiced by Make-10 Condition)								
Make 10	0.41	0.52	2.10	1.55	0.03	0.07	0.32	0.31
Near Doubles	0.64	0.75	1.22	1.12	0.05	0.11	0.14	0.21
Unpracticed Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Not Practiced in Any Condition)								
Make 10	0.45	0.61	1.42	1.23	0.05	0.09	0.19	0.23
Near Doubles	0.61	0.70	1.24	1.07	0.04	0.11	0.16	0.20
Practiced Near-doubles Combinations (Near-doubles Items Practiced by Near-doubles Condition)								
Make 10	0.75	0.89	1.48	1.39	0.09	0.17	0.20	0.27
Near Doubles	0.91	0.82	2.57	1.61	0.09	0.12	0.44	0.34
Unpracticed Near-doubles Combinations (Near-doubles Items Not Practiced in Any Condition)								
Make 10	2.22	1.88	2.84	1.57	0.39	0.39	0.48	0.35
Near Doubles	1.97	1.49	3.50	1.29	0.29	0.32	0.65	0.29
Practiced 7 + 8 and 8 + 7 (Items Practiced by both Conditions)								
Make 10	0.29	0.69	1.34	1.44	0.04	0.14	0.18	0.29
Near Doubles	0.36	0.54	2.00	1.93	0	0	0.36	0.41
Unpracticed 8 + 9, 9 + 8, and 9 + 9 (Items Not Practiced in Any Condition)								
Make 10	0.69	1.11	1.74	1.67	0.10	0.20	0.29	0.34
Near Doubles	0.56	0.88	1.83	1.30	0.06	0.15	0.27	0.24

Note. ^a Posttest scores adjusted for pretest score, free/reduced lunch eligibility, and age.

Table 8
Study 1 F-Index Posttest and (post – pretest difference) Performance Percentages

Mental Arithmetic Level						
	0	Phase 1 1	2	Phase 2 3	4	Phase 3 5
Practiced add-with-8 or -9 combinations						
Make-10	45.5 (-34)	9.6 (+1.5)	13.5 (+3.7)	0.9 (+0.9)	0 (0)	30.6 (+28)
Near-double	51.8 (-14.9)	11.5 (-4.3)	21.8 (+8.7)	0 (0)	0 (0)	14.9 (+10.4)
Transfer add-with-8 or -9 combinations						
Make-10	56.8 (-22.7)	7.9 (-1.9)	15.8 (+9.8)	0.4 (+0.4)	0 (0)	19.0 (+14.3)
Near-double	52.7 (-15.8)	11.3 (-2.7)	20.3 (+6.8)	0 (0)	0 (0)	15.8 (+11.7)
Practiced near-doubles combinations						
Make-10	54.1 (-16)	12.4 (+2.1)	13.9 (+2.8)	0 (0)	0 (0)	19.7 (+11.2)
Near-double	31.8 (-29)	5.9 (-9)	17.1 (+2.2)	0 (0)	0 (0)	45.3 (+35.8)
Transfer near-doubles combinations						
Make-10	26 (-18.2)	9.9 (+3.5)	13.5 (+3.2)	0 (0)	0 (0)	50.6 (+11.5)
Near-double	19.3 (-17.2)	4.4 (-12.5)	14.2 (-3.4)	0 (0)	0 (0)	62.2 (+33.1)

Table 9
Relation between Performance on Hypothesized Developmental Prerequisites and Mental-Addition Posttest Fluency of Near-Doubles items by Condition: Make-10 condition vs. Near-Doubles [brackets]

Developmental Prerequisite	Mental-addition posttest item			
	$2 + 3^1$		$3 + 2^1$	
$2 + 2^1$	Not	Fluent	Not	Fluent
Fluent	14 [13]	23 [21]	15 [11]	22 [23]
Not	1 [3]	1 [0]	1 [3]	1 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	$3 + 4$		$4 + 3$	
$3 + 3$	Not	Fluent	Not	Fluent
Fluent	26 [18]	8 [13]	22 [16]	12 [15]
Not	5 [6]	0 [0]	5 [4]	0 [2]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	$4 + 5^1$		$5 + 4^1$	
$4 + 4$	Not	Fluent	Not	Fluent
Fluent	26 [21]	9 [13]	24 [22]	11 [12]
Not	4 [3]	0 [0]	4 [3]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	$5 + 6$		$6 + 5$	
$5 + 5$	Not	Fluent	Not	Fluent
Fluent	31 [17]	7 [19]	31 [20]	7 [16]
Not	1 [1]	0 [0]	1 [1]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	$6 + 7$		$7 + 6$	
$6 + 6$	Not	Fluent	Not	Fluent
Fluent	22 [19]	2 [5]	24 [18]	0 [6]
Not	15 [13]	0 [0]	15 [13]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	$7 + 8$		$8 + 7$	
$7 + 7$	Not	Fluent	Not	Fluent
Fluent	9 [10]	4 [7]	11 [12]	2 [5]
Not	25 [17]	1 [3]	26 [17]	0 [3]
	$p = .011$ [$p = .046$]		$p < .001$ [$p = .018$]	
	$8 + 9^1$		$9 + 8^1$	
$8 + 8$	Not	Fluent	Not	Fluent
Fluent	13 [16]	4 [1]	12 [17]	5 [0]
Not	21 [20]	1 [0]	21 [19]	1 [1]
	$p < .001$ [$p < .001$]		$p = .002$ [$p < .001$]	

Note 1. "Fluent" for the developmental prerequisites indicates that a child met the criteria for fluency on one or both addition items; the "Not" indicates that a child did not meet fluency criteria on one or both prerequisites. A significant p value (McNemar test) confirms that a hypothesized prerequisite is a necessary condition for fluency with it.

Note 2. Groups appearing in **bold** practiced the prerequisite items in a given Frame.

Note 3. ¹ Indicates a transfer item.

Discussion

Discussed in turn are the implications of the results and limitations of the present study. All reports of fluency rate are limited to non-conscious reasoning strategies and automatic fact recall (5-point F-Index scores) as there were no fast, overt reasoning strategies (4-point scores) given during either testing session.

Implications of the Results

H1: Efficacy of Make-10 Training. H1 was not supported. The adjusted mean F-Index score—overall progress towards more efficient strategy use—for make-10 students on practiced items rose by +1.69 into the range of typically correct, but slow responses. The percentage decrease of make-10 students offering slow, counted, or incorrect responses was 32.5% versus 19.2% of near-doubles students; while the percentage of make-10 students using automatic or subconscious reasoning increased by 28% versus 10.4% of near-doubles students. With practiced add-with-8 or -9 items, the make-10 training resulted in significantly more Phase 3 responses than regular classroom instruction; the adjusted mean fluency rate (Phase 3) of make-10 students rose to one-third while merely one-seventh of near-doubles students attained that level.

On transfer items, both groups remained in the spectrum categorized by correct, but slow or counted responses. The percentage decrease of make-10 students offering slow, counted, or incorrect responses decreased by 24.6% versus 18.5% of near-doubles students; while the percentage of make-10 students using automatic or subconscious reasoning increased by 14.3% versus 11.7% of near-doubles students. Less than one-fourth of students in either training achieved Phase 3 on the posttest. Unfortunately, the guided instruction was not more effective than regular classroom instruction in promoting the learning of a general make-10 strategy necessary for transfer.

H2: Efficacy of Near-doubles Training. H2 was supported. On practiced and transfer near-doubles, the near-doubles training resulted in significant gains on both the F-Index and fluency rate measures over regular classroom instruction. On posttest practiced items, the near-doubles students had an adjusted F-Index score of 2.57, indicative of slow reasoning or deliberate undetermined strategies, a gain of 1.66 points, versus 1.48 (+0.73) for the make-10 students; while on transfer items, near-doubles students' score of 3.50, a remarkable gain of 1.53 points, was indicative of deliberate overt reasoning with no evidence of counting compared to a 2.84 (+0.62) score for make-10 students.

Furthermore, on the delayed posttest, over 33% of the near-doubles students improved their adjusted fluency rate (Phase 3) performance on both practiced and transfer items compared to an 11% improvement for make-10 students. This pattern of results is consistent with the conclusion that the gain in F-Index scores by the near-doubles group on practiced and transfer items was due to applying strategies more advanced than counting or slow, overt reasoning (Phases 1 and 2). The effect sizes well exceed the 0.25 criterion for substantively important practice set by the WWC handbook (IES, 2014).

H3: Developmental Prerequisites. The need for teaching of the highly salient doubles to foster the deduction of unknown doubles was supported. Knowing the developmental prerequisite doubles items is instrumental in having success solving near-doubles items.

Limitations

Although the findings are promising, some limitations must be noted. (a) Generalizability is limited to second graders having difficulties learning mathematics and results may not be representative of other categories of students. (b) The study lacks a training condition that involved practice-only in a semi-random order. No group practiced all targeted items from both

interventions therefore arguments cannot be made concerning the difference in structured discovery versus practice only. (c) There was a lack of an explanation for why an answer was either correct or incorrect for each trial. For beginning learners, knowing why an answer does or does not make sense could inform their ability to rationalize why a particular reasoning strategy is effective for a combination family. (d) The sample size per intervention was small. (e) Ideally, randomization should have occurred at the classroom level to negate the possibility of an effect due to differential instruction. (f) Given that this was an efficacy study and administered in a relatively structured research environment, standard classroom implementation is needed to gauge the true effectiveness of these interventions.

Conclusion

The Role of Guided Instruction. Consistent with the recommendations of NMAP (2008) and the number sense view: Structured discovery/practice can be an effective educational tool in promoting the learning of mathematical regularities and combination fluency especially as it relates to computer-assisted instruction (Clements & Sarama, 2012). Also, consistent with the NRC's recommendation (Kilpatrick et al., 2001), the learning of reasoning strategies can be furthered and often accelerated by conceptually based instruction.

The near-doubles training supported the recommendations of both advisory panels. Although the results regarding add-with-8 or -9 were not statistically significant, the effect sizes indicate that purposeful practice on these combinations is promising for grade 2 students having difficulties learning mathematics.

The Role of Practice Frequency. Practice as an instructional tool needs to be used purposefully and judiciously. The notion of hundreds or preferably thousands of repetitions as necessary to achieve (by rote) memorization as suggested by previous models and computer

simulations of arithmetic learning (Shrager & Siegler, 1998; Siegler & Jenkins, 1989) are misguided. Purposeful instruction involves recognizing how addition and subtraction are interdependent operations (Piaget, 1964). When students are developmentally ready, instruction should involve integrating the related operations of addition and subtraction instead of treating arithmetic fact families as independent associations. Practicing combinations before a child has mastered the developmental prerequisites for a combination family may be ineffective in promoting meaningful learning. Judiciously implies if structured correctly, students can have success learning the basic facts in much less than the recommended hundreds of repetitions.

Prior Knowledge and Salience. The complexity of the regularity (i.e., the difficulty of the steps involved) within a strategy may impact the learning of the targeted items. The ability to induce and assimilate the pattern or relation (i.e., salience) given a student's prior conceptual, factual, and procedural knowledge can be complicated by the level of fluency with the developmental prerequisite items. The near-doubles regularity seems to be highly salient for at-risk grade 2 students given their high aptitude with the doubles items and number-after knowledge.

Finally, future research would benefit from an add-with-8 or -9 training featuring two highly guided interventions, make-10, and use-ten (e.g., for the sum $9 + 7$: $10 + 7 = 17$ so $9 + 7$ is one less or 16). The goal, determining if either explicit, highly guided instruction is more effective than just practice and if there is a preference among students of make-10 versus use-ten as determined by their strategy usage rate. With the stated CCSSO (2.OA.4) standard "by the end of Grade 2, know from memory all sums of two 1-digit numbers" obtaining instructional insights regarding those students having difficulties learning mathematics would greatly benefit teachers and parents alike.

References

- Ackerman, P. T., Anhalt, J. M., & Dykman, R. A. (1986). Arithmetic automatization failure in children with attention and reading disorders: Associations and sequela. *Journal of Learning Disabilities, 19*, 222-232. doi:10.1177/002221948601900409
- Bajwa, N. P. (2013). *Evaluating the response time criterion for defining first graders' fluency with basic addition and subtraction* (Unpublished masters' thesis). University of Illinois, Champaign, IL.
- Baroody, A.J. (1999). Children's relational knowledge of addition and subtraction. *Cognition and Instruction, 17*, 137-175. doi:10.1207/S1532690XCI170201
- Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences, 6*, 1-36. doi:10.1016/1041-6080(94)90013-2
- Baroody, A. J. (1985). Mastery of the basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education, 16*, 83–98. doi:10.2307/748366
- Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review, 3*, 225-230. doi:10.1016/0273-2297(83)90031-X
- Baroody, A. J., & Purpura, D. J. (in preparation). Number and operations. In J. Cai (Ed.), *Mathematics education and research handbook*. Reston, VA: National Council of Teachers of Mathematics.
- Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Review, 15*, 69-79 (Special issue on "Pathways to Mathematical Learning Disabilities" guest edited by M. Mazzocco). doi:10.1002/ddrr.45

- Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development*, 20, 80-128. doi:10.1080/10409280802206619
- Baroody, A. J., Thompson, B., & Eiland, M. (2008). *Fostering the fact fluency of grade 1 at-risk children*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2014). *Fostering primary grade pupils' reasoning strategies with basic subtraction and relatively difficult addition combinations via computer-assisted instruction*. Manuscript submitted for publication.
- 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. doi:10.3102/0002831212473349
- 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. doi:10.1080/07370008.2012.720152
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300. doi:10.2307/2346101
- Brownell, W. A. (1935). Psychological considerations in the learning and the teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic* (Tenth yearbook, National Council of Teachers of Mathematics, pp. 1–31). New York: Bureau of Publications, Teachers College, Columbia University.
- Brownell, W. A. (1941). *Arithmetic in grades I and II: A critical summary of new and previously reported research*. Durham, NC: Duke University Press.

- Buckingham, B. R. (1927). Teaching addition and subtraction facts together or separately. *Educational Research Bulletin*, 6, 228-229, 240-242.
- Bullock, J. K. (1971-present). *Touch Math*. Colorado Springs, CO: Innovative Learning Concepts, Inc.
- 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. doi:10.1037/0012-1649.34.5.882
- Clements, D. H., & Sarama, J. (2012). Learning and teaching early and elementary mathematics. In Carlson, J., & Levin, J. (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. doi:10.1016/j.cogdev.2009.09.005
- Dowker, A. (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. 243-265). Mahwah, NJ: Erlbaum.
- Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*, 23, 45-55. doi:10.2307/749163

- Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction. *Cognition*, 123, 392–403. doi: 10.1016/j.cognition.2012.02.008
- Fuson, K. C. (2006). *Math expressions*. Boston, MA: Houghton Mifflin.
- Geary, D. C. (1996). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association. (Original work published 1994). doi:10.1037/10163-000
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, 33, 18–28.
doi:10.1177/002246699903300102
- Goldman, S. R., Pellegrino, J., & Mertz, D. L. (1988). Extended practice of basic addition facts: Strategy changes in learning disabled students. *Cognition and Instruction*, 5, 223–265.
doi:10.1207/s1532690xci0503_2
- Hatano, G. (2003). Forward. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. xi–xiii). Mahwah, NJ: Erlbaum.
- 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. doi:10.2307/30034895
- Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Erlbaum.
- Institute of Education Sciences (2014). *What Works Clearinghouse: Procedures and standards handbook (Version 3.0)*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf
- James, W. (1958). *Talks to teachers on psychology and to students on some of life's ideals*. New York, NY: W. W. Norton & Company. (Talk originally given in 1892.)
- Jerman, M. (1970). Some strategies for solving simple multiplication combinations. *Journal for Research in Mathematics Education*, 1, 95–128. doi:10.2307/748856
- Jordan, N. C. (2007). The need for number sense. *Educational Leadership*, 65(2), 63–66.
- Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, 85, 103–119. doi:10.1016/S0022-0965(03)00032-8
- Jordan, N. C., Hanich, L. B., & Uberti, H. Z. (2003). Mathematical thinking and learning difficulties. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 359–383). Mahwah, NJ: Erlbaum.
- Katona, G. (1967). *Organizing and memorizing: Studies in the psychology of learning and teaching*. New York, NY: Hafner. (Work originally published in 1940.)
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, D.C.: National Academy Press.
- Larson, N. (2008). *Saxon math 1*. Austin, TX: Harcourt Achieve

- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K. Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (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.
- Mason, J., & Spence, M. (1999). Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics*, 38, 135–161.
doi:10.1023/A:1003622804002
- Muarata, 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.
doi:10.1207/s1532690xci2202_2
- 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, D.C.: U.S. Department of Education.
- National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Washington, D.C.: National Academy Press.
- National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.
- Olander, H. T. (1931). Transfer of learning in simple addition and subtraction. *Elementary School Journal*, 31, 427–437. doi:10.1086/456594
- Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget*

- rediscovered* (pp. 7–20). Ithaca, NY: Cornell University.
- Purpura, D. J., Baroody, A. J., Eiland, M. D., & Reid, E. E. (2014). *Fostering reasoning strategies with the most basic sums via computer-assisted instruction: The value of guided training varies with combination family*. Manuscript submitted for publication.
- Rathmell, E. C. (1978). Using thinking strategies to teach basic facts. In M. N. Suydam & R. E. Reys (Eds.), *Developing computational skills* (1978 Yearbook, pp. 13–50). Reston, VA: National Council of Teachers of Mathematics.
- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93, 346–362. doi:10.1037/0022-0663.93.2.346
- Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction*, 1, 217–244. doi:10.1207/s1532690xci0102_3
- Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.
- Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, 9, 405–410. doi:10.1111/1467-9280.00076
- Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, 59, 833–851. doi:10.2307/1130252
- Siegler, R.S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.

- Skemp, R. R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher*, 26(3), 9–15.
- Skemp, R. R. (1979). *Intelligence, learning, and action*. Chichester, UK: Wiley.
- Skemp, R. R. (1987). *The psychology of learning mathematics*. Hillsdale, NJ: Erlbaum.
- Smith, J. H. (1921). Arithmetic combinations. *The Elementary School Journal*, 10, 762–770.
- Steinberg, R. M. (1985). Instruction on derived facts strategies in addition and subtraction. *Journal for Research in Mathematics Education*, 16, 337–355. doi:10.2307/749356
- Suydam, M., & Weaver, J. F. (1975). Research on mathematics learning. In J.N., Payne (Ed.), *Mathematics Learning in Early Childhood* (pp. 44–67). Reston, VA: NCTM.
- Swenson, E. J. (1949). Organization and generalization as factors in learning, transfer and retroactive inhibition. In *Learning theory in school situations* (No. 2). Minneapolis: University of Minnesota Press.
- Thiele, C. L. (1938). *The contribution of generalization to the learning of addition facts*. New York: Bureau of Publications, Teachers College, Columbia University.
- Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.
- Threlfall, J., Frobisher, L., & MacNamara, A. (1995). Inferring the use of recall in simple addition. *British Journal of Educational Psychology*, 65, 425–439. doi:10.1111/j.2044-8279.1995.tb01163.x
- 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. doi:10.1207/s1532690xci2301_1
- University of Chicago School Mathematics Project. (2005). *Everyday mathematics teacher's lesson guide (Volume 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.
- Wertheimer, M. (1959). *Productive thinking*. New York, NY: Harper & Row. (Original work published 1945.)
- Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604. doi:10.1037/0003-066X.54.8.

Chapter II.
FOSTERING ARITHMETIC FLUENCY WITH GRADE 1 AND GRADE 3 STUDENTS

ABSTRACT

Two training experiments involved investigating the efficacy of software designed to assist students in learning the more difficult early arithmetic combinations. In Experiment 1, advanced grade 1 students were randomly assigned to a moderately guided intervention featuring either a make-10 (e.g., the sum of $9 + 7$: $9 + [1 + 6] = [9 + 1] + 6 = 10 + 6 = 16$) or structured near-doubles (e.g., the sum of $3 + 4$: $3 + [3 + 1] = [3 + 3] + 1 = 6 + 1 = 7$) reasoning strategy. In Experiment 2, grade 3 students with difficulties learning mathematics were randomly assigned to one of three moderately guided interventions, the two outlined in experiment 1 or subtraction-as-addition (e.g., $12 - 9$ can be thought of as: What number when added to 9 equals 12?). ANCOVAs using age, pretest achievement, SES, and standardized mathematics achievement as covariates with mean delayed performance by combination family as the dependent variable were the units of analyses.

Computational fluency (appropriately, adaptively, and efficiently applying knowledge) on the basic number combinations has long been a goal of elementary mathematics instruction. Although there is disagreement as to how this goal is attained there is general agreement among mathematics educators that all children need to achieve this goal (e.g., Jordan, Hanich, & Kaplan, 2003; Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics or NCTM, 2000, 2006; National Mathematics Advisory Panel or NMAP, 2008). Due to the policy expectations of No Child Left Behind along with instructional benchmarks set forth by the Council of Chief State School Officers (CCSSO, 2010), fluency has received increased attention.

Although fluency with the basic sums and differences is one of the first benchmarks of primary education (CCSSO, 2010; Kilpatrick, et al., 2001; NCTM, 2000, 2006), many students struggle to achieve this goal (NMAP, 2008). For example, Henry and Brown (2008) documented first graders' performance on the California state standard of fluency using untimed tests of sums and differences and self-reports (both methods can exaggerate measures of fluency). They found that on testing batteries pertaining to sums and differences to 18 that median mastery was 22%. The results were representative of schools located in both low and high-performing districts despite an instructional emphasis on fluency on such items throughout the state. Children from low-income families or with limited instructional supports outside of the classroom may not achieve fluency with the basic combinations before the prescribed end of Grade 2 as stated by CCSSO (2010). A lack of fluency is a characteristic of those with difficulties learning mathematics (Ackerman, Anhalt, & Dykman, 1986; Geary, 1996; Goldman, Pellegrino, & Mertz, 1988; Jordan, Hanich, & Kaplan, 2003; Jordan, Hanich, & Uberti, 2003; Russell & Ginsburg, 1984).

Compounding the difficulties many children face in learning the basic combinations is a tradition of instruction that focuses on rote memorization and repeated practice. Such an approach is an outgrowth of Thorndike's (1922) *law of frequency* which promoted the notion that the more two stimuli are presented together (e.g., the more frequently a child sees an arithmetic combination such as "4 + 5" and the correct answer "9") the stronger the association between the two addends and the sum becomes—resulting in efficient recall of the correct answer when the arithmetic combination is presented. Certain associative models suggest hundreds or even thousands instances of practice are necessary to achieve efficient fact recall (Shrager & Siegler, 1998; Siegler & Jenkins, 1989). There are a number of drawbacks related to expending so much time and effort to memorize hundreds of basic combinations by rote but this instructional approach remains popular.

Promoting fluency can be more beneficial if achieved through meaningful memorization or learning, which entails linking conceptual, procedural, and factual knowledge (Hiebert & Carpenter, 1992; Hiebert & Lefevre, 1986; Katona, 1967; Mason & Spence, 1999; Moursund, 2002; Resnick & Ford, 1981; Rittle-Johnson, Siegler, & Alibali, 2001) such that students take advantage of patterns and relations inherent in addition and subtraction. Meaningful learning builds on students' existing knowledge and allows them to incorporate general principles and concepts when tackling moderately novel or new arithmetic situations. Most importantly, meaningful learning better promotes retention and transfer than does learning that does not make use of connections and sense-making regularities (Brownell, 1941; Brownell & Chazel, 1935; Gersten & Chard, 1999; Henry & Brown, 2008; James, 1958; Jordan, 2007; Katona, 1967; Kilpatrick et al., 2001; NMAP, 2008; Piaget, 1964; Skemp, 1978, 1979, 1987; Steinberg, 1985; Suydam & Weaver, 1975; Swenson, 1949; Thiele, 1938; Wertheimer, 1959).

The natural progression of meaningfully learning a basic combination or family of combinations typically involves three overlapping phases (Kilpatrick et al., 2001; Stienberg, 1985). In Phase 1 (counting) children use objects or verbal counting to determine sums. In Phase 2 (reasoning) children use known combinations and relations to consciously deduce the answer to an unknown sum. In Phase 3 (retrieval) children efficiently produce a sum or difference from memory. According to Brownell (1935), the “true test” of competency is “an intelligent grasp upon number relations and the ability to deal with arithmetical situations with proper comprehension of their mathematical as well as their practical significance (p. 19)”. Counting to determine a sum can help students develop patterns and relations that lead to reasoning strategies. With practice, reasoning strategies can become automatic (Jerman, 1970) and in time become mastered (Baroody, 1985). Baroody and Purpura (in preparation) rightly emphasize “there is no *a priori* reason to assume that reconstructive strategies cannot be compiled and become as automatic and reliable as a reproductive strategy and perhaps with significantly less time and effort (Baroody, 1985, 1994).” To become proficient with reasoning strategies, students need a strong number sense—an ever-developing network of meaningful number knowledge (Brownell, 1935; Gersten & Chard, 1999; Heavey, 2003; Jordan, 2007).

Number sense should build on what students already know (James, 1958; Piaget, 1964). For instance, Common Core grade 1 goals include subtraction-as-addition (Goal 1.OA.4) “understand subtraction as an unknown-addend problem. For example, subtract $10 - 8$ by finding the number that makes 10 when added to 8.” Reasoning strategies such as subtraction-as-addition were advocated in the early 1900s with Mead and Sears (1916) using Thorndike’s (1918) *identity of procedure* as justification posited “the ability developed by means of one subject can be transferred to another subject only in so far as the latter has elements in common with the former

(p. 277).” They found subtraction-as-addition is not suitable when drill is the preferred instructional method (Mead & Sears, 1916). Suzzalo (1911) stated the need for using relations more bluntly in a volume for teaching of primary arithmetic:

to learn two forms for one thing, is a waste... 'Subtracting by adding' is merely using the same association and word form for both addition and subtraction. Hence only one set of tables, instead of two, had to be learned. The meaning, the applicability, and the visual form of addition and subtraction are still different. Only the process of remembering and using the fundamental combinations is the same (p. 86-87).

Although the purposefulness of practice frequency plays a role, the importance related to the salience of the concepts or relations underlying the meaningful learning of a strategy cannot be ignored. Salience, the ability to induce and assimilate the pattern or relation given a students' prior conceptual, factual, and procedural knowledge can be complicated by the level of fluency with the developmental prerequisite items. For success using the subtraction-as-addition strategy, a child must first master the related addition items associated with the subtrahend and difference (e.g., for success with $13 - 9$, a student must be fluent with $9 + 4$).

Rationale for the Present Study

The purpose of the present studies was to evaluate the efficacy of experimental programs designed to foster a make-10 reasoning strategy, a near-doubles reasoning strategy, or a subtraction-as-addition reasoning strategy (experiment 2 only). Existing research suggests that basic differences are more difficult to learn than basic sums (Carpenter & Moser, 1984; Kraner, 1980; Smith, 1921; Woods, Resnick, & Groen, 1975; see Cowan, 2003, for a review). Although subtraction to twenty is a worthwhile early goal, educators often overlook the place value awareness needed regarding subtraction with regrouping. Gersten and Chard (1999) stress “[subtraction with regrouping] is the first math skill for which the child needs number sense to solve problems and, without such a sense, performance breaks down.” Teaching that subtraction

can be thought of as addition without regard for what the digits *one* and *six* represent in $16 - 8$ for example does little to enhance what the symbolism represents and what actions are required. A discussion of previous experimental efforts to foster fluency with larger sums, the reasons for the features of the present studies, and the research aims and hypotheses are discussed.

Previous Training Efforts

Add-With-8 or -9 Combinations. Previous efforts to teach a make-10 strategy (e.g., for the sum of $9 + 7$: $9 + [1 + 6] = [9 + 1] + 6 = 10$, $10 + 6 = 16$), even with supports to assist students in discovering the relationship themselves, failed to promote transfer with the strategy. Among students who have strong foundational knowledge of forming ten quickly, the make-10 strategy could be a powerful instructional tool. In Perry, VanderStoep, and Yu (1993), Perry cites her 1989 study in which first-grade students in Taiwan were taught addition and subtraction through decomposition and re-composition of addends using 10 as a placeholder. Murata (2004) cites similar instruction in Japanese first-grade classrooms with their break-apart-to-make-10 strategy (e.g., for $9 + 4$, think 9 and 1 make 10, separate 4 into 1 and 3, add 1 to 9 to make 10, add 3 more to get 13) which he describes as “a horizontal process of splitting one number into two parts” a distinction from the up-over-ten method reported by Fuson (1992) and others where students are instructed to go up to 10 and “up over” 10 (e.g., for $9 + 4$, $9 + 1$ then $10 + 3 = 13$).

For success with the make-10 strategy, a student must know: a) decomposition of the smaller addend such that the larger addend will sum to ten (fluency with decomposing numbers to 9); b) the remaining portion to retain of the decomposed smaller addend in working memory such that it can be added to ten; c) the $n + 10/10 + n$ facts (recognizing that any single-digit number n added to 10 results in the sum $n + \text{teen}$); and in certain situations d) the associative property of addition. Canobi et al. (1998) found that first and second graders reported using

associative-based relationships as a computational shortcut on only 11% of applicable problems taking on average just over 13 seconds to solve each problem. Two previous research efforts headed by Baroody (Baroody, Thompson, & Eiland, 2008; Baroody, Eiland, Paliwal, Priya-Bajwa, & Baroody, 2010) to promote a make-10 strategy failed to produce a significant transfer to unpracticed combinations.

Near-Doubles Combinations. Near-doubles training involved practice with $n + n$ items, $n + (n + 1) / (1 + n) + n$ items, and filler items. The general form is $n + m$, where $m = n + 1$, for example, $6 + 7 = 6 + (6 + 1) = (6 + 6) + 1 = 12 + 1 = 13$. For success, a student must know: the related double fact involving the smaller addend ($n + n$) and the next number in the count sequence following the double. Although it is equally accurate to have students solve $m + m - 1$, for the sake of simplicity, all near-doubles feedback was presented as $n + n + 1$ or $1 + n + n$.

The sixth goal in the operations and algebraic thinking domain for grade 1 Common Core State Standard (1.OA.6) states: “add and subtract to 20, demonstrating fluency for addition and subtraction within 10. Use such strategies as...making ten (e.g., $8 + 6 = 8 + 2 + 4 = 10 + 4 = 14$) and creating equivalent but easier known sums (e.g., adding $6 + 7$ by creating the known equivalent $6 + 6 + 1 = 12 + 1 = 13$).” Teachers often invoke direct instruction of reasoning strategies. Imposing such reasoning strategies often does not work (Torbeyns, Verschaffel, & Ghesquire, 2005) because children have difficulty determining appropriate use—selectively applying knowledge to suitable cases accurately (e.g., the $7 + 6$ is $6 + 6 + 1$ or $7 + 7 - 1$ is 13) but other times inaccurately (e.g., for $7 + 6$: misuse on $6 + 6 - 1$ is 11 or that $7 + 7 + 1$ is 15). Therefore, efforts to impose reasoning strategies on children, particularly those at risk for academic failure, may be ineffective when children do not understand the rationale of the imposed procedure. Previous research efforts had mixed results in promoting fluency with a

near-doubles strategy, Baroody et al. (2010) was successful in promoting significant transfer to unpracticed items with grade 2 students while Baroody, Thompson, and Eiland (2008) was unsuccessful with advanced grade 1 students.

Experiment 1 – Grade 1 Students

Features of the Current Experimental Training

General Program Improvements. A concentrated effort was made to improve the level of student engagement with the computer interventions. Students in grade 1 are capable of understanding how mathematical ideas come to be known and understood (Perry, McConnery, Flevares, Mingle, & Hamm, 2011) when empowered to do so. A number of improvements were made in relation to the Year 3, Fall, 2007 study:

1. A 5-stage approach was adopted. In stage I, students are asked to identify the symbolic representation for a word problem. Students are allowed to solve the problem using any method of their choosing.
2. In stage II, students are charged with identifying the proper range for a quantity of size 1 to 20. In a comparable investigation, students are asked to identify the proper answer range for a posed addition or subtraction item. Participants in Year 1 to Year 3 of the project were asked “if you do not know the answer, give a good guess or an answer that makes sense.” After three years of testing hundreds of students, the determination was made that students with a weak number sense have difficulty providing a “good guess” or answers within 25% of a sum in the correct direction consistently. This estimation training was developed to assist students in deciding what constitutes a “good guess”.
3. Stages III to V are specific to an intervention and introduce students to a targeted reasoning strategy and progressively wean them from concrete solution to solutions

generated from recall.

4. A new game, *Does It Help?*, explicitly asks students if an initial combination can be used to deduce the answer to a follow-up combination.

A seven week training experiment with advanced grade 1 students served to evaluate the efficacy of two experimental programs for fostering reasoning strategies. (The Institute of Educational Sciences' definition and requirements for efficacy can be found at http://ies.ed.gov/funding/pdf/2012_84305A.pdf.) At the beginning of the school year, students were introduced to strategies that would help them: estimate quantities and sums, and learn to relate adding-1 to the next number in the count sequence. Students who mastered the add-1 relation and the doubles relation by the March posttest were randomly assigned to either a make-10 or a near-doubles intervention for the remainder of the school year.

Hypotheses

The following three hypotheses were evaluated:

Hypothesis 1: The make-10 training should facilitate the meaningful learning of a general make-10 addition reasoning strategy above and beyond regular classroom instruction received by its active control (the near-doubles group). At the delayed posttest, the make-10 group should have significantly better mean fluency index (F-Index) and mean fluency rate performance for the practiced and unpracticed add-with-8 or -9 items than the active control group.

Hypothesis 2: The near-doubles training should facilitate the meaningful learning of a general near-doubles addition reasoning strategy above and beyond regular classroom instruction as represented by the make-10 group. At the delayed posttest, the near-doubles group should have significantly better mean fluency index (F-Index) and mean fluency rate performance for the practiced and unpracticed $n+(n+1)/(1+n)+n$ items than the make-10 group.

Hypothesis 3: Success with the doubles items should lend itself to success with the near-doubles items. At the delayed posttest, is success with the doubles items, as determined by the fluency rate measure, a necessary condition for success on the near-doubles items?

Methods

Participants

Participants were recruited from 15 grade 1 classrooms in five elementary schools from adjacent school districts serving two medium-sized mid-western cities during Fall, 2008.

Parental consent forms were returned for 122 students. At the end of March, 45 students did not know at least 67% of the targeted add-with-1 items and doubles items and thus continued to study those combinations until the end of the school year. Seventy-seven students knew more than 67% of add-with-1 and doubles items and were pretested on add-with-8 or -9 and near-doubles items. Of the 77 students, 3 were successful on more than 75% of the targeted items and were not included in the study (too advanced). All 74 eligible students completed the study focusing on add-with-8 or -9 or near-doubles with no attrition.

Among the 74 participating students (6.0 to 7.5 years of age; mean = 6.6 years; median = 6.6 years old; SD = 0.3 years), 54.1% were male. The composition of the sample was 45.9% Caucasian; 35.1% African-American; 8.1% Multi-ethnic, 4.1% Hispanic; 2.7% Asian, and 4.1% Other, respectively. Additionally, 39.2% of participants were eligible for free or reduced-price lunch. 16.2% of participants had no at-risk factor (i.e., eligible for free/reduced lunch, TEMA-3 achievement score ≤ 90 , teacher identified low achievement, English as a Second Language, or documented medical condition). Participant demographic information by condition is summarized in Table 10.

Table 10
Experiment 1 Participant Characteristics by Condition

Characteristic	Training Condition	
	Structured Make-10	Structured Near-doubles
Age range	6.0 to 7.5	6.0 to 7.0
Mean (SD)	6.6 (0.4)	6.6 (0.3)
Median age	6.6	6.6
School		
1	3	3
2	7	8
3	10	10
4	12	12
5	4	5
Boys (Girls)	17 (19)	23 (15)
TEMA-3 range	78 to 127	72 to 125
Mean (SD)	99.2 (12.3)	98.6 (9.8)
Median	98.5	98.5
Students $\leq 25^{\text{th}}$ percentile on TEMA-3	8	10
Free/Reduced lunch eligible	16	13
Black/Hispanic/Multiracial	16	19
English as Second Language	2	4
Medical Condition		
ADHD	1	2
Other	0	1
Teacher Identified	14	12
No Risk Factor	6	6

All five participating schools were committed to achieving the State's grade 1 learning objectives that included operations on whole numbers such as solving one- and two-step problems and performing computational procedures using addition and subtraction <<http://www.isbe.state.il.us/ils/math/capd.htm>>. The ten classes in Schools 1, 2, and 4 used *Everyday Mathematics* (University of Chicago School Mathematics Project or UCSMP, 2005). The five classes in Schools 3 and 5 used *Math Expressions* (Fuson, 2006); teachers in School 3 supplemented computation practice with the *Touch Math* (Innovative Learning Concepts Inc., 2011) program. See Appendix E for information concerning how these curricula treat adding-with-8 or -9 and near-doubles.

Project hired personnel consisted of three female Academic Professionals (APs) and nine female and one male Research Assistants (RAs). Eight staff members (All three APs and five RAs) also served on Study 1. Prior to the beginning of the study all staff members had six 3-hour training sessions on testing/training procedures.

Administrative Procedure. Testing and training was done on a one-to-one basis with a project staff member typically in 30-minute sessions, twice a week. Testing and training were conducted at project-dedicated workstations (new Apple iMac desktops with trainer-operated touchscreen technology) in a designated research space (e.g., shared reading room or auxiliary classroom) or a hallway outside the students' classroom. The new desktops allowed the computer to read a majority of the task instructions to students; the touchscreen equipment allowed the trainer to control program functions and input detailed students' mental-arithmetic responses. Students who knew at least 75% of the add-with-1 and doubles items on the March posttest were randomly assigned at the school level to either the make-10 intervention or the near-doubles intervention with groups balanced by performance on all add-with-8 or -9 items. A delayed posttest, identical to the pretest was administered at least two weeks after the conclusion of the training.

Measures

TEMA-3 testing. Students in Experiment 1 were administered a mathematics achievement test during the first 3 weeks of September. The Test of Early Mathematics Achievement-Third Edition (TEMA-3; Ginsburg & Baroody, 2003), a manually and individually administered, nationally standardized games-based test of mathematics achievement for children aged 3-years to 8-years 11 months, was used to gauge the baseline knowledge of students by establishing a standardized achievement score. The test measures general number sense as well

as concepts in the domains of: fluency with number combinations, calculation skills, and understanding of arithmetic concepts. Cronbach's alphas for (a) 7 year olds is .95 for the testing form used (Form A); (b) males, females, European, African, Hispanic, and Asian Americans are all .98; and (c) low mathematics achievers, .99. In terms of criterion-predictive validity, correlations between the TEMA-3 and similar measures (the KeyMath-R/NU, Woodcock-Johnson III, Diagnostic Achievement Battery, and Young Children's Achievement Test) range from .54 to .91. The TEMA-3 administration was handled by a specially trained group of four RAs headed by the author. All scoresheets were reviewed for scoring accuracy. Scaled scores of ≤ 90 points (25th percentile or less) were used as an indicator of "at-risk".

Mental-arithmetic Test. The preliminary mental-arithmetic pretest, featuring adding-with-1 and doubles, occurred in the last 2 weeks of March. The mental-arithmetic test featuring adding-with-8 or -9, near-doubles, and filler items occurred immediately after identifying students who knew at 75% of the add-1 items. The seven categories of problems appearing on the pre/posttest are summarized in Table 11:

Table 11
Experiment 1 Tested Items by Type and Condition

Combination Family	Combinations	Practiced in Stages III to V by Group	
		Structured Make-10	Structured Near-doubles
Practiced add-with-8 or -9 items	$3 + 9, 6 + 9, 7 + 9, 9 + 3, 9 + 6$	Yes	No
Transfer add-with-8 or -9 items	$4 + 9, 5 + 9, 9 + 4, 9 + 5, 9 + 7$	No	No
Practiced plus-10 items	$2 + 10, 7 + 10, 10 + 4, 10 + 6, 10 + 8$	Yes	No
Practiced items for both groups	$5 + 10, 9 + 8, 10 + 3$	Yes	Yes
Transfer items	$8 + 9, 9 + 9$	No	No
Practiced doubles items	$3 + 3, 4 + 4, 5 + 5, 6 + 6, 7 + 7, 8 + 8$	No	Yes
Practiced Near-doubles items	$3 + 4, 4 + 3, 5 + 6, 6 + 5, 7 + 8$	No	Yes
Transfer doubles and near-doubles items	$4 + 5, 5 + 4, 6 + 7, 7 + 6, 8 + 7$	No	No
Practiced filler items	$3 + 5, 4 + 7, 5 + 3, 7 + 4$	5 + 3 No	4 + 7 & 5 + 3 Yes 3 + 5 & 7 + 4 No

Note. The mental-arithmetic pretest/posttest was composed of four sets of 10 items each. Set 1 in order was: $5 + 5, 10 + 4, 7 + 9, 5 + 4, 9 + 3, 6 + 7, 4 + 3, 8 + 8, 5 + 10$ and $4 + 9$. Set 2 in order was: $10 + 8, 6 + 5, 8 + 9, 3 + 3, 5 + 9, 2 + 10, 9 + 6, 7 + 8, 5 + 3$, and $9 + 7$. Set 3 in order was: $7 + 10, 9 + 8, 7 + 6, 9 + 4, 4 + 7, 10 + 6, 3 + 9, 4 + 5, 7 + 7$, and $3 + 4$. Set 4 in order was: $5 + 6, 8 + 7, 3 + 5, 9 + 9, 10 + 3, 6 + 9, 4 + 4, 9 + 5, 6 + 6$, and $7 + 4$.

Training Intervention

Training consisted of 12 sets divided into three Stages and lasted for 7 weeks. Each set was divided into two subsets. A complete overview of the aim and the plan for each set can be found in Table 12. New curricular emphases in Experiment 1 included: a) 16 weeks of preparatory work (Stages I and II) common to all students that assisted in translating word problems into symbolic representations, as well as an estimation task that asked students to guess which range of numbers best fits a collection ranging between 0 to 20 items or which range of numbers is the best guess for an addition expression; b) in the *Does It Help?* Game, students are shown a problem (plus-ten/doubles) and asked if it will assist in solving the next problem; and c)

in the revised *Castle Wall Game*, students click on the wall to input their response to the prompted problem, new decomposition prompts regarding how many dots need to be removed to complete the targeted reasoning strategy, and finally an equation enters the wall summary, $\boxed{3 + 4} = \boxed{7}$, instead of the less explicit symbolism $\boxed{3 + 4 \rightarrow 7}$ that was used as summary feedback in Year 3. The *Does it Help? Game* explicitly relates the targeted strategy with the developmental prerequisite. Comparison between the models is required by the child. The *Castle Wall Game* implicitly relates the targeted strategy with the developmental prerequisite. Comparison between the models is not required by the child.

Table 12
Experiment 1 and Experiment 2 Mental-Arithmetic Stages

Preliminary Stages			
Stage	Computer Game	Rationale	
I	Word Problems	Determine which number sentence represents the stated word problem.	
II	Tammy the Frog	Determine which of the four ranges contains the best estimate of a collection of 1 to 20 frogs.	
	About Where in the World	Determine which of three or four ranges contains the best estimate for an addition or a subtraction item.	
	Billy the Goat A	Determine which of the four ranges contains the best estimate of a collection of 1 to 20 carrots.	
	Billy the Goat B	Determine which of four ranges contains the best estimate for an addition or a subtraction item.	
Targeted Interventions			
Stage/Set	Computer Game	Structured Make-10	Structured Near-doubles
III/A	Castle Wall	Solve add-with-10 item such as $10 + 7$, followed by a related add-with-8 or -9 item such as $9 + 8$; or solve add-with-8 or -9 item followed by its commuted partner.	Solve a doubles item such as $7 + 7$, followed by solving a related near-doubles item such as $7 + 8$.

Table 12 continued

III/B	Does It Help? (Possible helper and target items presented successively. Helper problems have the same sum as target items.)	Solve an $n + 10/10 + n$ item, such as $5 + 10$; then asked if it helps solve an $n + 8/8 + n$ or $n + 9/9 + n$ item such as $9 + 6$ (yes) or $9 + 8$ (no).	Solve a doubles item such as $3 + 3$; then asked if it helps to solve $2n + 1$ or $n + m$ such as $3 + 4$ (yes) or $4 + 7$ (no).
IV/A	Castle Wall / Does It Help?	Same as above	Same as above
IV/B	Number Line Jumble (Possible helper and target items occasionally presented successively. Sums appear haphazardly underneath a number line missing all but the larger addend)	Solve an $n + 10/10 + n$ item, such as $10 + 7$ followed by a complementary $n + 9/9 + n$ item (e.g., $9 + 8$) or a decoy problem (e.g., $7 + 4$).	Solve a doubles item such as $5 + 5$ by locating the sum among the jumbled elements below the number line; then asked to solve related near-doubles item (i.e., $5 + 6$).
V/A V/B	Mental arithmetic preparation (Immediate feedback)	Trained items including add-with-8, -9, or -10 and filler items.	Trained items including doubles, near-doubles, and filler items.

In Stage III (Sets 30-33), problems are presented concretely and symbolically. Each problem is depicted concretely and the targeted strategy is actively modeled. All sets are untimed with feedback provided at the end of a subset based on 10-stars. The developmental prerequisite (plus-10 or doubles) item always preceded the targeted (add-with-8 or -9 or near-doubles) item. The first subset was the *Castle Wall Game*. Using $3 + 4$ as an example, the student first enters her/his answer by clicking on the appropriate 1 to 20 wall block. The computer prompts “How many do we need to remove from here (highlighted 4) to make the doubles three and three more? Click on as many dots as you need to change four into three.” The revised problem $3 + 3 + 1$ is displayed along with “Removing only one dot changed four into a double”. Students are encouraged to solve the doubles portion using the 5-frame model “Use the five frames to figure out how much three and three more is. Click on the castle blocks to answer”. After answering,

the two frames containing three each combine and with the prompt “The answer is 6.” Finally, the student is asked “How much is six and one more altogether? Enter your answer on the number under the wall.” After clicking on their response, the computer displays “The answer is 7”, a single red dot enters the ten-frame with the other red dots, $6 + 1 = 7$. A block displaying $3 + 4 = 7$ entered the castle wall.

The second subset involved the *Does it Help? Game*. Students are shown a plus-10 or a doubles problem and asked if it will assist in solving the next problem. Using the paired example $5 + 10 / 6 + 9$, $5 + 10$ is shown and then its sum “Five and ten more is fifteen. Does five and ten more is fifteen help you to answer what six and nine more is?” A number line representation from 1 to 20 is displayed with a number bar of length 5 adjacent to a bar of length 10 above the number line. Two choices are displayed at the top of the screen: ☐ No, because six and nine is not fifteen or ☐ Yes, because six and nine is also fifteen. After the child selects her/his answer, the computer automatically arranges a number bar of length 6 and another of length 9 above the bars of length 5 and 10 such that the make-10 relation is more readily understood. The computer displays “Let’s see if five and ten more helps to answer six and nine more. If six and nine more is also fifteen.” The numeral 6 is highlighted, the numeral 9 is highlighted and while the two bars are combined, the prompt “Six and nine more is fifteen” is displayed. After the two number bars combine, the prompt “So, yes, five and ten more is fifteen helps to answer what six and nine more is because six and nine more is also fifteen.” After 10 problem pairs have been completed, the computer displays feedback based on 10-stars.

Stage IV (Sets 34-37) of the training was similar to Stage III with the following exceptions:

- a) Items were initially presented as symbolic expressions; mental solutions were encouraged for the first try. Concrete modeling of a targeted strategy occurred as in Stage III if there was an incorrect response on the first try.
- b) Targeted items sometimes follow the related developmental prerequisite.
- c) Targeted reasoning strategy is not highlighted. Equations or concrete models of related items never presented simultaneously.
- d) Time component added. For Sets 34-35, speed = 25%, accuracy = 75%; Sets 36-37, speed = 33%, accuracy = 67%.

Stage V (Sets 38-41) mimics the mental-arithmetic test. An item is displayed symbolically in the center of the screen which triggers the clock. After the student responds and as the tester enters the answer, the correct answer, response time, and feedback out of 5-stars is displayed. In Sets 38-39, feedback is based on speed = 33% and accuracy = 67%; Sets 40-41, present feedback using the same criteria as the mental-arithmetic test: speed = 50% and accuracy = 50%. See Appendix F for screen shots of Stages III to V.

All students saw the correct answer the same number of times. Targeted items in general were practiced no more than 16 times (except for developmental prerequisites $3 + 3$ and $5 + 5$ which were practiced 25 times) during the 7 weeks of training.

Role of the Project Trainers. Administrative duties included completing lesson and mental-arithmetic test log sheets to ensure each participant completed each lesson or testing session without duplication. Logistical duties included picking up children from their classrooms, obtaining positive assent for each testing and training session, escorting children to a project workstation, logging them into the appropriate training program, and encouraging on-task behavior. The trainer's primary instructional role was to give voice to the scripted instructions,

read number sentences during mental-arithmetic testing, and explain feedback graphically displayed by the computer screen.

Fidelity of Intervention. Fidelity of training was ensured by (a) 10 hours of staff training prior to the beginning of the project on the rationale for the programs and procedures for implementing the computer guided training, (b) a copy of the Trainer Guidelines at each computer station, (c) brief (10 to 30 minute) staff meetings during the semester to review procedures and address training issues as needed, and (d) a lesson log sheet for keeping track of which lessons each student had completed. The programs ensured that each child received (a) the assigned intervention (upon logging in the child was automatically connected to his or her treatment), (b) the combinations in the order specified by an intervention, and (c) feedback on correctness. In regard to implementation fidelity, all students saw the correct answer the same number of times and completed all 20 sets before the posttest. Each targeted item was practiced 20 times during the 9 weeks of training.

The make-10 and near-doubles groups practiced mutually exclusive items therefore, served as an active control for each other. Various threats to internal validity were accounted for by the random assignment. Significant posttest differences cannot be attributed to history (e.g., classroom instruction or practice), regression to the mean, maturation, or selection, because theoretically random assignment ensures all groups are comparable on these confounding variables. Students received identical tests the same number of times regardless of training condition to discount a testing effect. Both groups received identical *Castle Wall Game* and *Does It Help? Game* training interfaces as well as identical reward games via the computer to control for a novelty effect. Any contamination or diffusions effect, which would facilitate the learning

of the near-doubles and add-with-8 or -9 items by sensitizing the child to the mathematical regularities adds to measurement error and makes it more difficult to obtain significant results.

Analytic Procedure

ANCOVAs, using experimental condition as the grouping variable and age, free/reduced lunch eligibility, mental-arithmetic pretest scores, and TEMA-3 achievement pretest score as the covariates, were used to compare posttest performance of each group on targeted practiced and transfer combinations. A correction (Benjamini & Hochberg, 1995) was applied to adjust for Type I error due to multiple comparisons. The adjustment was applied separately for the near-doubles and add-with-8 or -9 items. For each, there were a total of two comparisons (practiced items and transfer items). Additionally, effect size magnitude (Hedges' g) was examined for all contrasts due to the limited power of the study and the importance of evaluating effect sizes (Wilkinson & APA Task Force on Statistical Inference, 1999). Effect sizes were calculated after accounting for the covariates using the posttest mean proportion correct. Per the Institute of Education Science's (IES) What Works Clearinghouse (WWC) "effects that are not statistically significant but have an effect size of at least 0.25 are considered 'substantively important'" (IES, 2014, p. 23). A Hedges' g that exceeds 0.25 will serve as the effect size measure to evaluate whether a particular intervention is substantively important as well as feasible. In cases where Levene's Test of Equality of Error Variances was violated, a condition*pretest interaction was added into the model. Gender, school, and classroom were initially included in the analysis as random variables. Since none of these variables were statistically significant, they were removed from all analyses. Thus all reported analyses feature experimental condition as the only random variable. Random assignment to treatment occurred at the school level. Gender, school, and classroom were initially included in the analysis as random variables. For the analyses involving

8+9/9+9 and 9+8 there was statistically significant school effect, therefore, school is included as a random variable in those analyses. Except for the 8+9/9+9 and 9+9 analyses, all other reported analyses feature experimental condition as the only random variable.

Ten students on the pretest and five students on the posttest consistently created a teen using one of the addends. The mean plus-10 combination performance for these students was adjusted to account for the false positives (64 items total). A complete summary of false positive performance can be found in Appendix G.

A significant p value on McNemar's test was used to confirm whether or not prerequisite knowledge (plus-10/doubles) is necessary for success on targeted items (add-with-8 or -9 and near-doubles).

Results

Preliminary results and the delayed posttest results regarding Hypothesis 1 to Hypothesis 2 and the developmental prerequisite analysis regarding Hypothesis 3 are discussed. All reported statistically significant results for main effects of condition remained so when the Benjamini-Hochberg adjustment for multiple comparisons was applied. Given the directional nature of the contrasts, effects of treatment were tested using 1-tailed significance values unless noted.

Pretest Analyses

Pretest analyses revealed that the composition of children in the two groups did not differ in gender, $\chi^2(1, N=74) = 1.32, p = .251$; free/reduced lunch eligibility, $\chi^2(1, N=74) = 0.81, p = .367$, or ethnicity (comparing African American, Asian, Caucasian, Latino/a, Multi-racial, and other), $\chi^2(5, N=74) = 0.77, p = .979$. There were no differences among groups in terms of age, $F(1, 72) = 0.05, p = .818$, or on the TEMA-3, $F(1, 72) = 0.06, p = .804$. Using proportion correct, 2-tailed ANOVAs revealed no significant differences among groups for the targeted add-with-8

or -9, $F(1, 72) = 0.01$, $p = .962$, and near-doubles, $F(1, 72) = 0.01$, $p = .981$, on and all items respectively. Cronbach's alpha for all add-with-8 or -9 items and all near-doubles items were $\alpha = .34$ and $\alpha = .75$, respectively. See Table 13 and Table 14 for student performance results.

Hypothesis 1: Efficacy of the Structured Make-10 Training

F-Index. For add-with-8 or -9 items, after applying the Benjamini-Hochberg adjustment, there was a marginally significant performance difference on practiced items, $F(1, 67) = 3.06$, $p = .043$, $g = 0.92$, but no difference on the transfer items, $F(1, 67) = 0.84$, $p = .182$, $g = 0.50$. On items where either reasoning strategy was applicable, there were no statistically significant differences for, either the practiced $8 + 9/9 + 9$ items, $F(1, 64) = 2.35$, $p = .135$ (2-tailed), $g = 0.30$, or the $9 + 8$ transfer item, $F(1, 64) = 0.42$, $p = .521$ (2-tailed), $g = 0.15$ were significant. In the case of all items except $9 + 8$, the effect sizes exceed the IES (2014) Hedges' $g = 0.25$ criteria for substantively important practice favoring the make-10 intervention.

Fluency Rate. On items scored as fluent on the F-Index¹, after applying the Benjamini-Hochberg adjustment, the structured make-10 group significantly outperformed the near-doubles group on practiced add-with-8 or -9 items, $F(1, 67) = 12.27$, $p < .001$, $g = 1.01$, and more importantly, on transfer items, $F(1, 67) = 7.24$, $p = .005$, $g = 0.67$. The effect sizes for both practiced and transfer items exceed the IES (2014) $g = 0.25$ criteria for substantively important practice favoring the make-10 intervention. On items where either reasoning strategy was applicable, there were no significant differences for, either the practiced items $8 + 9/9 + 9$, $F(1, 68) = 1.16$, $p = .286$ (2-tailed), $g = 0.22$, or the $9 + 8$ transfer item, $F(1, 68) = 0.33$, $p = .567$ (2-tailed), $g = 0.13$.

¹ There were no 4-point scores on either the pre or the post mental-arithmetic test.

Table 13
Experiment 1 Pretest and Adjusted Posttest F-Index and Fluency Rate Scores by Condition

Condition	F-Index (0 to 5)				Fluency rate (0 to 1)			
	Pretest		Adjusted ^a Posttest		Pretest		Adjusted ^a Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Practiced Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Practiced by Make-10 Condition)								
Make 10	1.03	0.73	2.45	1.36	0.08	0.15	0.36	0.30
Near Doubles	0.97	0.71	1.38	0.91	0.08	0.13	0.12	0.16
Unpracticed Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Not Practiced in Any Condition)								
Make 10	0.57	0.65	1.69	1.23	0.01	0.05	0.19	0.21
Near Doubles	0.70	0.70	1.15	0.87	0.03	0.08	0.07	0.14
Practiced Near-doubles Combinations (Near-doubles Items Practiced by Near-doubles Condition)								
Make 10	1.42	1.23	1.65	1.17	0.17	0.25	0.21	0.24
Near Doubles	1.46	1.05	2.01	1.31	0.16	0.22	0.27	0.29
Unpracticed Near-doubles Combinations (Near-doubles Items Not Practiced in Any Condition)								
Make 10	0.87	0.89	1.42	1.13	0.09	0.18	0.18	0.23
Near Doubles	0.99	0.82	1.32	0.97	0.10	0.15	0.16	0.19
Practiced 8 + 9 and 9 + 9 (Items Practiced by both Conditions)								
Make 10	0.65	1.02	2.19	1.71	0.10	0.20	0.27	0.35
Near Doubles	0.96	1.11	1.70	1.54	0.09	0.20	0.20	0.27
Unpracticed 9 + 8 (Item Not Practiced in Any Condition)								
Make 10	0.58	1.27	2.36	2.13	0.06	0.23	0.34	0.49
Near Doubles	0.39	0.75	2.05	1.96	0	0	0.28	0.45

Note. ^a Posttest scores adjusted for age, free/reduced lunch eligibility, pretest score, and TEMA-3 standardized score.

Table 14

Experiment 1 – F-Index Posttest and (post – pretest difference) Performance Percentages

Mental Arithmetic Level						
	0	Phase 1 1	2	Phase 2 3	4	Phase 3 5
Practiced add-with-8 or -9 combinations						
Make-10	25 (-30)	9.4 (-1.7)	29.4 (+3.8)	0.6 (0)	0 (0)	35.6 (+27.8)
Near-double	43.7 (-14.2)	11.6 (0)	32.6 (+10.5)	0.5 (0)	0 (0)	11.6 (+3.7)
Transfer add-with-8 or -9 combinations						
Make-10	50 (-16.1)	12.2 (-2.2)	28.9 (+10.6)	0 (0)	0 (0)	8.9 (+7.8)
Near-double	45.3 (-15.2)	24.2 (+7.4)	25.3 (+5.8)	0 (-0.5)	0 (0)	5.3 (+2.7)
Practiced near-doubles combinations						
Make-10	41.7 (-8.9)	16.1 (+7.8)	21.1 (-2.8)	0 (0)	0 (0)	21.1 (+3.9)
Near-double	32.1 (-12.6)	14.2 (+0.5)	26.3 (+1)	0.5 (+0.5)	0 (0)	26.8 (+10.5)
Transfer near-doubles combinations						
Make-10	50 (-13.3)	15.6 (+2.8)	17.2 (+2.2)	0 (0)	0 (0)	17.2 (+8.3)
Near-double	48.9 (-10.6)	13.7 (+1.6)	21.1 (+2.7)	0 (0)	0 (0)	16.3 (+6.3)

Hypothesis 2: Efficacy of the Structured Near-Doubles Training

F-Index. For near-doubles items, after applying the Benjamini-Hochberg adjustment, the near-doubles group marginally outperformed the make-10 group on practiced items, $F(1, 68) = 2.73, p = .052, g = 0.29$, with no differences on transfer items, $F(1, 67) = 0.49, p = .244, g = -0.10$. The effect size for the practiced items exceeded the IES (2014) $g = 0.25$ criteria for effective practice favoring the near-doubles intervention.

Fluency Rate. After applying the Benjamini-Hochberg adjustment, the results for the structured near-doubles group versus the structured make-10 group on near-doubles items indicated no differences in performance on practiced, $F(1, 67) = 0.97, p = .164, g = 0.22$, or transfer items, $F(1, 68) = 0.29, p = .297, g = -0.10$.

Hypothesis 3: Developmental Prerequisite Knowledge

Regardless of intervention, on all twelve comparisons, students who were not fluent on prerequisite plus-10 items were not fluent on the corresponding add-with-8 or -9 items for both practiced and transfer items on the posttest. Students in the near-doubles group, on all twelve comparisons, who were not fluent on the prerequisite doubles item were not fluent on the corresponding practiced and transfer near-doubles items. The same was true for students in the make-10 training except in two instances, $4 + 4/5 + 4$ ($p = .072$) and $8 + 8/8 + 9$ ($p = .194$); also more students were fluent on $8 + 9$ than with $8 + 8$. Student performance can be found in Table 15.

Table 15

Relation between Performance on Hypothesized Developmental Prerequisites and Mental-Addition Posttest Fluency of Make-10 items (Frame A) and Near-Doubles items (Frame B) by Condition: Make-10 condition vs. Near-Doubles [brackets].

Developmental Prerequisite	Mental-addition posttest item			
Frame A: Plus-ten items x adding-with-8 or -9 items				
	5 + 9		9 + 5	
5 + 10	Not	Fluent	Not	Fluent
Fluent	32 [33]	3 [1]	27 [33]	8 [1]
Not	1 [3]	0 [1]	1 [4]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	7 + 9		9 + 7 ¹	
7 + 10	Not	Fluent	Not	Fluent
Fluent	32 [33]	4 [2]	28 [30]	16 [9]
Not	0 [3]	0 [0]	0 [3]	2 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	3 + 9		9 + 3	
10 + 3	Not	Fluent	Not	Fluent
Fluent	16 [27]	17 [8]	17 [26]	16 [9]
Not	2 [3]	1 [0]	1 [3]	2 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	4 + 9 ¹		9 + 4 ¹	
10 + 4	Not	Fluent	Not	Fluent
Fluent	31 [31]	5 [3]	31 [20]	7 [16]
Not	0 [4]	0 [0]	1 [1]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	6 + 9		9 + 6	
10 + 6	Not	Fluent	Not	Fluent
Fluent	23 [31]	11 [1]	22 [31]	12 [1]
Not	2 [6]	0 [0]	1 [5]	1 [1]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	8 + 9 ¹		9 + 8	
10 + 8	Not	Fluent	Not	Fluent
Fluent	25 [31]	10 [5]	22 [26]	13 [10]
Not	1 [2]	0 [0]	1 [2]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	

Table 15 continued

<i>Frame B: Doubles items x near-doubles items</i>				
	3 + 4		4 + 3	
3 + 3	Not	Fluent	Not	Fluent
Fluent	25 [21]	5 [14]	20 [25]	10 [10]
Not	6 [3]	0 [0]	6 [2]	0 [1]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	4 + 5 ¹		5 + 4 ¹	
4 + 4	Not	Fluent	Not	Fluent
Fluent	17 [16]	6 [15]	12 [18]	11 [13]
Not	11 [7]	2 [0]	8 [6]	5 [1]
	$p < .001$ [$p < .001$]		$p = .072$ [$p < .001$]	
	5 + 6		6 + 5	
5 + 5	Not	Fluent	Not	Fluent
Fluent	22 [23]	12 [15]	23 [27]	11 [11]
Not	2 [0]	0 [0]	2 [0]	0 [0]
	$p < .001$ [$p < .001$]		$p < .001$ [$p < .001$]	
	6 + 7 ¹		7 + 6 ¹	
6 + 6	Not	Fluent	Not	Fluent
Fluent	9 [21]	2 [1]	9 [22]	2 [0]
Not	25 [16]	0 [0]	24 [16]	1 [0]
	$p = .002$ [$p < .001$]		$p = .011$ [$p < .001$]	
	7 + 8		8 + 7 ¹	
7 + 7	Not	Fluent	Not	Fluent
Fluent	10 [17]	0 [0]	9 [17]	1 [0]
Not	26 [21]	0 [0]	25 [20]	1 [1]
	$p = .001$ [$p < .001$]		$p = .011$ [$p < .001$]	
	8 + 9 ¹		9 + 8	
8 + 8	Not	Fluent	Not	Fluent
Fluent	4 [19]	2 [4]	3 [14]	3 [9]
Not	22 [14]	8 [1]	20 [14]	10 [1]
	$p = .194$ [$p < .001$]		$p = .046$ [$p < .001$]	

Note 1. "Fluent" for the developmental prerequisites indicates that a child met the criteria for fluency on one or both addition items; the "Not" indicates that a child did not meet fluency criteria on one or both prerequisites.

Note 2. Groups appearing in **bold** practiced the prerequisite items in a given Frame.

Note 3. ¹ Indicates a transfer item.

Discussion

Discussed in turn are the implications of the results and limitations of the present study.

All reports of fluency rate are limited to non-conscious reasoning strategies and automatic fact recall (5-point F-Index scores) as there were no fast, overt reasoning strategies (4-point scores) given during either testing session.

Implications

H1: Efficacy of Make-10 Training. The results indicate that implicit and explicit training on the make-10 reasoning strategy is efficacious for advanced Grade 1 students. The adjusted mean F-Index score for make-10 students on practiced items rose +1.42 into the range of typically correct, but slowly reasoned responses while near-doubles students rose only +0.41 into the range of slow undetermined strategies or counted responses. The percentage decrease of make-10 students offering slow, counted, or incorrect responses was 31.7% versus 14.2% of near-doubles students; whereas the percentage of make-10 students using automatic or subconscious reasoning increased by 27.8% versus 3.7% of near-doubles students. With practiced add-with-8 or -9 items, both groups were similar in terms of producing deliberate reasoning strategies (Phase 2) but the make-10 training resulted in significantly more Phase 3 responses than regular classroom instruction; the adjusted mean fluency rate (Phase 3) of make-10 students rose to over one-third while less than one-eighth of near-doubles students reached that level of proficiency.

On transfer items, both groups were in the spectrum categorized by correct, but slow or counted responses. The percentage decrease of make-10 students offering slow, counted, or incorrect responses decreased by 18.3% versus 7.8% of near-doubles students; while the percentage of make-10 students using automatic or subconscious reasoning increased by 7.8% versus 2.7% of near-doubles students. Less than one-fifth of students in either training achieved Phase 3 on the posttest. At first glance the results may not seem worthwhile; however, there was a significant performance difference by make-10 students on 5-point scoring (Phase 3) transfer items.

H2: Efficacy of Near-doubles Training. The adjusted mean F-Index score for near-doubles students on practiced items rose by +0.55 into the range of typically correct, but slow responses while the make-10 students rose +0.32 and stayed in the range of slow undetermined strategies or counted responses. The percentage decrease of near-doubles students offering slow, counted, or incorrect responses was 12.1% versus 1.1% for make-10 students; while the percentage of near-doubles students using automatic or subconscious reasoning increased by 10.5% versus 3.9% for make-10 students. With practiced near-doubles items, there were no differences in Phase 3 performance between groups (adjusted mean fluency rate improvement: near-doubles group .11; make-10 group .04).

The results for transfer near-doubles items were unexpected, favoring the make-10 students. The percentage decrease of near-doubles students offering slow, counted, or incorrect responses was 9.0% versus 10.5% for make-10 students; while the percentage of near-doubles students using automatic or subconscious reasoning increased by 6.3% versus 8.3% for make-10 students. The near-doubles students were outperformed on both the F-Index and fluency rate measures by the make-10 students although there were no statistically significant differences on either measure. With transfer near-doubles items, there were no differences in Phase 3 performance between groups (adjusted mean fluency rate improvement: near-doubles group .06; make-10 group .09).

H3: Developmental Prerequisites. The teaching of the highly salient doubles to foster the deduction of unknown doubles and the teaching of the highly salient plus-ten items to foster the deduction of unknown add-with-8 or -9 items was supported. Knowing the developmental prerequisite doubles and plus-ten items is instrumental in having success solving near-doubles and add-with-8 or -9 items.

Experiment 2 – Grade 3 Students

Features of the Current Experimental Training

General Program Improvements. A new subtraction-as-addition intervention was developed to assist students in the learning of the subtraction combinations by using their addition knowledge.

1. In Stage II, the Billy the Goat game interface was simplified. Students are now presented with two options, **Smart Guess** or **Silly Guess**. If silly guess is chosen, a second submenu of answer options was presented.
2. A new subtraction-as-addition intervention introduced students to the part-part-whole notion of fact triads.

Subtraction-as-Addition Program. For Piaget (1965), a thorough understanding of number entailed understanding *additive composition*—realizing that addition and subtraction are interdependent operations. To effectively learn the subtraction-as-addition strategy meaningfully such an understanding is paramount. As summarized by Baroody, Purpura, Eiland, and Reid (2014) additive composition involves three key relations:

- Complement principle and combination families. Algebraically, the *addition-subtraction complement principle* is represented as: If $a + b = c$ or $b + a = c$, then $c - b = a$ or $c - a = b$. This principle is the basis for the subtraction-as-addition reasoning strategy (Common Core Goal 1.OA.4: understand subtraction as an unknown-addend problem) and implies that certain addition and subtraction combinations (e.g., $8 + 6 = 14$, $6 + 8 = 14$, $14 - 8 = 6$, and $14 - 6 = 8$) belong to the same “family.”
- Inversion. The immediate recognition that adding a number b to a number a can be undone by subtracting the same number b and vice versa ($a + b - b$ or $a - b + b = a$).

Baroody, Torbeyns, and Verschaffel (2009) note that recognizing that adding b to a can be undone by taking away b may serve as a conceptual bridge for the complement principle if $a + b = c$, then $c - b = a$. Canobi (2004) found that children who understood the inverse principle with high levels of proficiency also knew the complement principle.

- Part-part-whole relations. According to Piaget and others (Briars & Larkin, 1984; Canobi, 2005; Resnick, 1983; Riley, Greeno, & Heller, 1983), the complement principle, the subtraction-as-addition strategy, and the inverse principle are an outgrowth of part-whole knowledge. In essence, if Part 1 + Part 2 equals the Whole ($P_1 + P_2 = W$), then taking a part from the whole should leave the other part. This relationship is the basis behind the “Math Mountains” featured in *Math Expressions* (Fuson, 2006) and the “fact triangles” featured in *Everyday Mathematics* (University of Chicago School Mathematics Project or UCSMP, 2005).

An example of what the moderately guided subtraction training entails is summarized in Table 16. A one semester training experiment with grade 3 students served to evaluate the efficacy of three experimental programs for fostering reasoning strategies.

Hypotheses

The first two hypotheses, for make-10 and near-doubles, are identical to those in Experiment 1; the third hypothesis is related to the new subtraction-as-addition intervention. Hypothesis 3: The subtraction-as-addition training should facilitate the meaningful learning of a general subtraction reasoning strategy above and beyond regular classroom instruction as represented by the make-10 and near-doubles groups. At the delayed posttest, the subtraction group should have significantly better mean F-Index measure and better mean fluency rate measure performance for the practiced and unpracticed subtraction items than the make-10 and

near-doubles groups. Finally, Hypothesis 4 is related to Hypothesis 3 in Experiment 1 regarding whether or not there is a need for fluency on developmental prerequisites (plus-10, doubles, and addition complement items) in order to have fluency on the targeted (add-with-8 or -9, near-doubles, and subtraction) items.

Table 16

A Continuum of Approaches for Teaching Reasoning Strategies [And How a Computer Program Might Embody Each Approach]

Feature	Unguided discovery	Partially guided				Fully guided
		Negligibly guided discovery	Minimally guided discovery	Moderately guided discovery	Highly guided discovery	
Structure + explicit scaffolding						
Strategy and its rationale (relations) explicitly taught Program explicitly specifies and illustrates the subtraction-as-addition rule: To find the missing value, state the number that when added to the subtrahend equals the minuend.	No	No	No	No	No	Yes
Explicitly underscore relation/strategy Subtraction training includes games that explicitly ask if a particular addition item helps to solve a particular subtraction problem.	No	No	No	No	Yes	Yes
Structure that provides implicit scaffolding						
Structured practice: related items sequenced to underscore the relation/strategy Subtraction items immediately follow related addition complement item (the developmental prerequisite for subtraction-as-addition) increasing the chances a child will note the connection between subtraction and existing knowledge of addition relations.	No	No	No	Yes	Yes	Yes
Implicitly underscore relation/strategy Child enters difference of subtraction items by clicking on a number list, which implicitly underscores addition complement relations.	No	No	Yes	Yes	Yes	Yes
No structure or scaffolding Subtraction items presented randomly; no related addition complements practice provided. No scaffolding other than feedback. Child enters a difference by clicking on a virtual keypad (a display that underscores subtraction complements relations to lesser extent than a number list.)						
Program-chosen (prescribed) games Child progresses through a sequence of pre-planned games that involve subtraction.	No	Yes	Yes	Yes	Yes	Yes
Child-selected (no prescribed) games Child chooses from a menu of addition and non-arithmetic games.	Yes	No	No	No	No	No

Methods

Participants

Participants were recruited from 10 classrooms in three elementary schools from adjacent school districts serving two medium-sized mid-western cities during Fall, 2009. Parental consent forms were returned for 119 students. Six students did not participate in the study because they either moved prior to participant assignments ($n = 3$), did not complete mental-arithmetic pretest ($n = 1$), joined the study after random assignment ($n = 1$), or because of grade reassignment ($n = 1$)². Study participants met the following criteria: fluent on less than 50% of the subtraction and add-with-8 or -9 items on the pretest. Pretest results indicated 14 students knew $\geq 50\%$ of the subtraction and add-with-8 or -9 (tested out of the study); following assignment to an intervention, another four students were lost due to moving ($n = 3$) or refusing to participate ($n = 1$). Among the 95 participating students who completed the study (8.0 to 9.5 years of age; mean = 8.5 years & median = 8.6 years old; SD = 0.37), 49.5% were male and 88.4% had at least one risk factor using the same guidelines outlined in Study 2. The composition of the sample was 44.2% African-American; 33.7% Caucasian; 9.5% Hispanic; and 12.6% multi-ethnic, Asian, unknown, or other race children, respectively. Additionally, 72.6% of the participants were eligible for free or reduced-price lunch. This demographic was chosen because of the high likelihood they would not be fluent with the basic sums and differences and might benefit from the interventions. See Table 17 for the demographic information by condition.

² The reassigned student was placed in a study involving grade 2 students the very next semester.

Table 17
Study 3 Participant Characteristics by Condition

Characteristic	Training Condition		
	Structured Make-10	Structured Near-doubles	Structured Subtraction
Age range	8.0 to 9.5	8.0 to 9.3	8.0 to 9.5
Mean (SD)	8.6 (0.4)	8.5 (0.3)	8.5 (0.4)
Median age	8.6	8.6	8.5
Boys (Girls)	16 (16)	16 (15)	15 (17)
School			
1	9	10	9
2	9	11	10
3	14	10	13
TEMA-3 range	70 to 124	74 to 127	67 to 130
Mean (SD)	90.4 (14.8)	93.3 (14.8)	93.1 (17.9)
Median	84.5	88	89.5
Students $\leq 25^{\text{th}}$ percentile on TEMA-3	22	17	18
Free/Reduced lunch eligible	24	23	22
Black/Hispanic/Multiracial	22	20	18
English as Second Language	0	2	2
No Risk Factor	1	5	6
Attrition	1 moved	1 refused	1 moved

All three participating schools were committed to achieving the State's grade 3 learning objectives that included operations on whole numbers such as solving one- and two-step problems and performing computational procedures using addition and subtraction <<http://www.isbe.net/ils/math/capd.htm>>. The seven classes in Schools 1 and 3 used *Everyday Mathematics* (University of Chicago School Mathematics Project or UCSMP, 2005). The three classes in School 2 used *Math Expressions* (Fuson, 2006) while supplementing computation practice with *Touch Math* and also incorporating materials from a Houghton Mifflin basal. Details regarding how these two curricula approached arithmetic, in general, and subtraction and add-with-8 or -9 items, in particular, can be found in Appendix E.

Project hired personnel consisted of seven female Academic Professionals (APs) and one male and eight female Research Assistants (RAs). Four of the APs and six of the RAs had previous experience on the project (0 to 5 years, median = 2 years and 0 to 5 years median = 2 years, respectively). Prior to the beginning of the study all staff members had six 3-hour training sessions: on testing/training procedures; strategies for promoting a positive disposition among participants; as well as strategies for both resolving and avoiding conflict.

Measures

TEMA-3 testing was conducted concurrently with Stage I training. The mental-arithmetic pretest was administered upon the conclusion of Stage II. The ten categories of problems appearing on the pre/posttest are summarized in Table 18:

Table 18
Study 3 Tested Items by Combination Category

Combination Family	Combinations	Practiced in Stages III to V by Group		
		Structured Make-10	Structured Near-doubles	Structured Subtraction
Practiced add-with-8 or -9	$6 + 9, 7 + 9, 8 + 5, 9 + 4, 9 + 6, 9 + 7$	Yes	No	$9 + 4$ Yes
Transfer add-with-8 or -9	$5 + 8, 5 + 9, 6 + 8, 8 + 6, 9 + 5$	No	No	No
Practiced add-with-10	$3 + 10, 4 + 10, 6 + 10, 7 + 10, 10 + 4, 10 + 7, 10 + 8^a$	Yes	No	No
Practiced Near-doubles	$3 + 4^a, 4 + 3, 5 + 4, 5 + 6, 6 + 5$	No	Yes	No
Transfer near-doubles	$4 + 5, 6 + 7, 7 + 6, 8 + 9, 9 + 8$	No	No	No
Practiced doubles	$7 + 7, 8 + 8$	No	Yes	$7 + 7$ Yes
Practiced subtraction	$10 - 5, 10 - 7, 11 - 7, 12 - 6, 12 - 8, 12 - 9, 13 - 4, 14 - 7$	No	No	Yes
Transfer subtraction	$10 - 6, 11 - 4, 11 - 8, 12 - 5, 12 - 7, 13 - 9$	No	No	No
Practiced addition complements	$4 + 7, 7 + 5$	No	No	Yes
Practiced filler items	$5 + 7, 7 + 4$	Yes	Yes	Yes

Note 1. ^aItem $3 + 4$ replaced item $10 + 8$ on the posttest.

Note 2. Prior to the mental-arithmetic test, students were given a pretest containing items of interest in order to introduce them to test expectations.

Set 0. $4 + 8, 5 + 10, 6 + 6, 11 - 6, 4 + 9, 10 - 8, 4 + 4, 3 + 9, 4 + 3, 13 - 7, 5 + 5$, and $10 - 6$.

The mental-arithmetic pretest/posttest consisted of four sets of 12 items each.

Set 1 in order was: $3 + 4$ ($10 + 8$), $12 - 9, 6 + 8, 5 + 7, 9 + 4, 4 + 5, 7 + 10, 5 + 9, 10 - 5, 9 + 6, 11 - 8$, and $8 + 9$.

Set 2 in order was: $7 + 7, 6 + 5, 14 - 7, 5 + 8, 4 + 10, 9 + 8, 12 - 5, 6 + 9, 7 + 4, 12 - 7, 8 + 8$, and $13 - 4$.

Set 3 in order was: $9 + 5, 10 - 6, 7 + 9, 10 + 4, 7 + 6, 12 - 8, 4 + 7, 6 + 10, 8 + 5, 7 + 8, 11 - 7$, and $5 + 6$.

Set 4 in order was: $11 - 4, 8 + 6, 3 + 10, 8 + 7, 13 - 9, 10 + 7, 5 + 4, 9 + 7, 10 - 7, 7 + 5, 12 - 6$, and $6 + 7$.

Training Intervention

Equipment description and performance indicators are the same as those outlined in

Experiment 1. Students who knew less than 50% of the add-with-8 or -9 and subtraction items

on the pretest were randomly assigned at the classroom level to one of three interventions: make-

10, near-doubles, or subtraction-as-addition.

The entire training was composed of 25 sets (consisting of two subsets each) separated into five stages and lasted for 9 weeks. Stages I and II were common to all participants and served to prepare students for the computer-assisted mental-arithmetic testing and the experimental interventions. The purpose was to insure that all students had the necessary experiences to benefit from the computerized programs (e.g., how to use the virtual manipulatives and enter answers) and the developmental prerequisites (e.g., solving word problems concretely and realizing that a good estimate for a subtraction problem must be smaller than both the minuend and subtrahend) to benefit from the primary training. One minor change from Experiment 1, the initial choices for Billy the Goat's guess now read **Silly Guess** and **Smart Guess**. If **Silly Guess** is chosen, a submenu of **Silly Small** or **Silly Big** appears.

The last three stages of the training were the structured interventions. The training in all three interventions was done in the context of computer games, with each session consisting of 20 to 24 items. Stage III (Sets 9-14) used the *Does It Help? Game* and *Castle Wall Game* described in Study 2. The aim of this Stage was for children to work through a reasoning strategy using a concrete model with relations highlighted both implicitly (*Castle Wall Game*) and explicitly (*Does It Help? Game*). Stage IV (Sets 15-20) serves as a transition from concrete to mental strategies. Initially, students see symbolic representations only, ten-frames or number lines were used as a backup for incorrect solutions. Feedback was generated on varying weights of accuracy and reaction time: sets 15 and 16 were 75% accuracy and 25% reaction time, sets 17 to 20 were 67% accuracy and 33% reaction time. Stage V (Sets 21-25) was symbolic only, omitting concrete aids. It was designed to mimic the expectations of the mental-arithmetic posttest. Students were encouraged to answer as quickly as possible "If you do not know an answer make a smart guess as fast as you can." Following the mental-arithmetic test guidelines,

accuracy and speed were equally weighted for the purposes of feedback. All students saw the correct answer the same number of times. No targeted item was practiced more than 24 times during the 9 weeks of training. For illustrations and descriptions of Stages I and V see Appendix H.

Fidelity of Intervention. Fidelity of training was ensured by (a) a copy of the Trainer Guidelines at each computer station, (b) brief (10 to 30 minute) staff meetings during the semester to review procedures and address training issues as needed, (c) a lesson log sheet for keeping track of which lessons each student had completed, and (d) the design of the computer games was such that each participant received only their assigned curriculum during Stages III to V. All participants completed the TEMA-3 and 100% of lessons in Stages I & II prior to the pretest and intervention assignment and Stages III to V prior to the delayed posttest.

Various threats to internal validity were accounted for by the random assignment at the classroom level. Significant posttest differences cannot be attributed to history (e.g., classroom instruction or practice), regression to the mean, maturation, or selection, because theoretically random assignment ensures all groups are comparable on these confounding variables. Students received identical tests the same number of times regardless of training intervention to discount a testing effect. All groups received training and reward games via the computer to control for a novelty effect. Any contamination or diffusions effect, which would facilitate the learning of add-with-8 or -9, near-doubles and subtraction items by sensitizing the child to the mathematical regularities, adds to measurement error and makes it more difficult to obtain significant results.

Analytic Procedure

ANCOVAs, using experimental condition as the grouping variable and age, free/reduced lunch eligibility, pretest scores, and TEMA-3 achievement score as the covariates, were used to

compare posttest performance of each group on targeted practiced and transfer combinations. The main intervention group was compared to the combined active control groups for each of the three primary analysis sets (add-with-8 or -9, near-doubles, and subtraction). A correction (Benjamini & Hochberg, 1995) was applied to adjust for Type I error due to multiple comparisons. Additionally, effect size magnitude (Hedges' g) was examined for all contrasts due to the limited power of the study and the importance of evaluating effect sizes (Wilkinson & APA Task Force on Statistical Inference, 1999). Effect sizes were calculated after accounting for the covariates using the posttest mean proportion correct.

On the pretest, five students consistently stated the subtrahend and had their subtraction combination mean performance adjusted to account for the false positives (five items total). On the posttest, two students stated a favorite number on subtraction items during Session 1; another student consistently stated the subtrahend. The mean subtraction combination performance for these students was adjusted to account for the false positives (3 items total). A complete summary of false positive performance can be found in Appendix I.

Although participants were randomly assigned within classrooms, analyses were also conducted with school and classroom included as random-effect covariates to check for school and classroom effects. Classroom effects were present in four contrasts and thus those analyses contain classroom random variable: a) practiced near-doubles items on the F-Index measure; b) practiced add-with-8 or -9 items, practiced make-10, and $8+9/9+8$ on the fluency rate measure. In cases where Levene's Test of Equality of Error Variances was violated, a condition*pretest interaction was added into the model. Except in the cases previously noted, neither school nor classroom is included in the final analyses.

A significant p value using McNemar's test was used to confirm whether or not prerequisite knowledge (plus-10/doubles/addition complements) is necessary for success on targeted (add-with-8 or -9, near-doubles, and subtraction) items.

Results

Pretest results

Pretest analyses revealed that the composition of children in the three groups did not differ in gender, $\chi^2(2, N=95) = 0.15, p = .929$; or ethnicity (comparing African American, Asian, Caucasian, Hispanic, Multi-racial, and other), $\chi^2(10, N=95) = 6.29, p = .791$. Using nationally normed TEMA-3 standardized scores, an ANOVA did not reveal significant group differences in mathematics achievement, $F(2, 92) = 0.32, p = .726$. Using mean proportion correct, ANOVAs revealed no significant differences (2-tailed) among groups for the targeted adding-with-8 or -9, $F(2, 92) = 1.06, p = .352$, near-doubles $F(2, 92) = 1.20, p = .305$, and subtraction combinations $F(2, 92) = 0.14, p = .871$ on all items. Cronbach's alpha for all add-with-8 or -9 items, all near-doubles items, and all subtraction items were $\alpha = .77$, $\alpha = .67$, and $\alpha = .74$, respectively.

See Table 19 and Table 20 for student performance results.

Hypothesis 1: The Efficacy of the Make-10 Training

F-Index. Planned comparisons for the structured make-10 group versus the active control groups, who received instruction on either structured near-doubles or structured subtraction-as-addition, on practiced add-with-8 or -9 indicated no statistically significant differences in performance on practiced items, $F(1, 33.656) = 0.30, p = .293, g = 0.56$, but a marginally significant difference on the transfer items, $F(1, 88) = 2.06, p = .078, g = 0.26$. For items that are similar to the make-10 and the near-doubles interventions, there was a significant difference on practiced items $7+8/8+7$, $F(1, 88) = 9.22, p = .002$ (2-tailed), $g = 0.68$, favoring the make-10

group, but not transfer items $8+9/9+8$, $F(1, 18.562) = 0.42$, $p = .369$ (2-tailed), $g = 0.13$. The effect sizes for practiced items, transfer items, and items practiced by both groups exceed the IES (2014) $g = 0.25$ criteria for substantively important practice favoring the make-10 intervention.

Fluency Rate. A planned comparison indicated both school and classroom effects on practiced add-with-8 or -9 items. Since random assignment was conducted at the school level, classroom was included as a random effect in this analysis. For add-with-8 or -9 items, there was no difference in performance between the structured make-10 group and the active-control groups, $F(1, 21.717) = 0.22$, $p = .321$, $g = 0.67$, or transfer items, $F(1, 88) = 1.54$, $p = .109$, $g = 0.25$. Although neither the practiced nor transfer results were statistically significant, the effect size for both comparisons exceeds the IES (2014) $g = 0.25$ criterion for substantively important practice favoring the make-10 intervention. For items that are similar to the make-10 and near-doubles interventions, there is a significant difference between the combined make-10 and near-doubles group on practiced items $7+8/8+7$, $F(1, 88) = 9.06$, $p = .002$, $g = 0.60$, but not on transfer items $8+9/9+8$, $F(1, 89) = 0.11$, $p = .374$, $g = 0.06$.

Hypothesis 2: The Efficacy of the Near-Doubles Training

F-Index. For near-doubles items, there was difference in performance between the structured near-doubles group and the active controls (structured make-10 and structured subtraction) groups on practiced items, $F(1, 33.901) = 0.39$, $p = .285$, $g = 0.33$, or transfer items, $F(1, 4.548) = 0.63$, $p = .233$, $g = 0.17$. Although neither the practiced nor transfer results are statistically significant, the effect size for the practice items comparison exceeded the IES (2014) $g = 0.25$ criterion for substantively important practice.

Fluency Rate. There was a marginal difference in performance between the structured near-doubles group and the active controls, $F(1, 89) = 2.99$, $p = .044$, $g = 0.28$, but no difference

on transfer items, $F(1, 34.799) = 0.83$, $p = .184$, $g = 0.28$. Although neither the practiced nor transfer results are statistically significant, the effect size for both comparisons exceed the IES (2014) $g = 0.25$ criterion for substantively important practice.

Table 19
Experiment 2 Pretest and Adjusted Posttest F-Index and Fluency Rate Scores by Condition

Condition	F-Index (0 to 5)				Fluency rate (0 to 1)			
	Pretest		Adjusted ^a Posttest		Pretest		Adjusted ^a Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Practiced Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Practiced by Make-10 Condition)								
Make-10	1.02	0.82	2.28	1.48	0.07	0.12	0.35	0.34
Near-doubles	1.40	1.17	1.52 ^b	1.38	0.15	0.19	0.16 ^b	0.27
Subtraction	1.13	1.23	1.53 ^b	1.21	0.11	0.20	0.16 ^b	0.23
Unpracticed Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Not Practiced in Any Condition)								
Make-10	1.13	0.93	1.67	1.05	0.11	0.17	0.21	0.22
Near-doubles	1.33	1.32	1.24 ^b	1.38	0.15	0.25	0.14 ^b	0.25
Subtraction	1.09	1.13	1.37 ^b	1.33	0.11	0.20	0.15 ^b	0.24
Practiced Near-doubles Combinations (Near-doubles Items Practiced by Near-doubles Condition)								
Make-10	2.23	1.74	2.46 ^c	1.42	0.35	0.35	0.37 ^c	0.32
Near-doubles	2.30	1.93	2.86	1.66	0.35	0.43	0.45	0.40
Subtraction	1.74	1.67	2.32 ^c	1.52	0.28	0.32	0.34 ^c	0.33
Unpracticed Near-doubles Combinations (Near-doubles Items Not Practiced in Any Condition)								
Make-10	1.91	1.30	1.64 ^c	1.26	0.28	0.28	0.16 ^c	0.25
Near-doubles	1.47	1.16	1.98	1.40	0.16	0.21	0.27	0.30
Subtraction	1.41	1.19	1.74 ^c	1.63	0.16	0.21	0.21 ^c	0.30
Practiced Subtraction Combinations (Subtraction Items Practiced by Subtraction Condition)								
Make-10	1.26	0.93	1.66 ^d	0.97	0.17	0.18	0.23 ^d	0.19
Near-doubles	1.53	1.09	1.64 ^d	1.30	0.21	0.21	0.23 ^d	0.24
Subtraction	1.31	1.26	2.24	1.23	0.18	0.22	0.34	0.26
Unpracticed Subtraction Combinations (Subtraction Items Not Practiced in Any Condition)								
Make-10	0.88	0.96	1.07 ^d	1.08	0.11	0.17	0.10 ^d	0.19
Near-doubles	0.83	0.89	1.13 ^d	1.24	0.08	0.14	0.11 ^d	0.19
Subtraction	0.73	0.85	1.47	1.23	0.07	0.13	0.16	0.22

Note. ^aAdjusted for covariates: age, free/reduced lunch eligibility, pretest score, and TEMA-3 achievement score. ^bAdjusted means from contrast with make-10 students. ^cAdjusted means from contrast with near-doubles students. ^dAdjusted means from contrast with subtraction students.

Table 20
Experiment 2 F-Index Posttest and (post – pretest difference) Performance Percentages

Mental Arithmetic Level						
	0	Phase 1	2	Phase 2	4	Phase 3
		1		3		5
Practiced Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Practiced by Make-10 Condition)						
Make-10	30.7 (-22.9)	14.1 (+1.6)	27.6 (+1.6)	1 (0)	0.5 (0)	26 (+19.2)
Near-doubles	37.6 (-8.1)	17.7 (+3.2)	24.2 (+2.7)	2.7 (-0.5)	0 (0)	17.7 (+2.6)
Subtraction	39.6 (-13)	13 (-2.6)	31.8 (+12.5)	0 (-1.6)	0 (0)	15.6 (+4.7)
Transfer Add-with-8 or -9 Combinations (Add-with-8 or -9 Items Not Practiced by Any Condition)						
Make-10	43.8 (-11.8)	13.1 (+3.7)	21.3 (-2.5)	1.9 (+1.9)	0.6 (+0.6)	19.4 (+8.1)
Near-doubles	48.4 (+0.7)	17.4 (-0.7)	16.8 (-1.3)	1.9 (+0.6)	0 (0)	15.5 (+0.7)
Subtraction	49.4 (-5.6)	8.1 (-3.8)	26.3 (+3.8)	0.6 (+0.6)	0 (-0.6)	15.6 (+5.6)
Practiced Near-doubles Combinations (Near-doubles Items Practiced by Near-doubles Condition)						
Make-10	28.9 (-9.6)	10.2 (+1.9)	23.4 (+7.8)	1.6 (-0.5)	0.8 (+0.8)	35.2 (-0.2)
Near-doubles	19.4 (-9.6)	12.1 (-6.2)	18.5 (+3.4)	4 (+0.8)	0 (0)	46 (+11.6)
Subtraction	36.7 (-13.3)	7 (-3.4)	21.9 (+11.5)	1.6 (+0.6)	0 (-1)	32.8 (+5.7)
Transfer Near-doubles Combinations (Near-doubles Items Not Practiced by Any Condition)						
Make-10	38.5 (-2.1)	11.5 (-3.1)	29.2 (+14.6)	2.1 (0)	0 (0)	18.8 (-9.3)
Near-doubles	35.5 (-8.6)	10.8 (-4.3)	26.9 (+4.3)	2.2 (-1)	0 (0)	24.7 (+9.6)
Subtraction	38.5 (-7.3)	15.6 (0)	20.8 (-1.1)	4.2 (+3.2)	0 (0)	20.8 (+5.2)
Practiced Subtraction Combinations (Subtraction Items Practiced by Subtraction Condition)						
Make-10	46.1 (-11.7)	11.7 (+1.5)	19.9 (+5.4)	0.8 (+0.4)	0 (0)	21.5 (+4.3)
Near-doubles	47.2 (-0.4)	10.1 (-4.8)	17.7 (+2)	1.2 (+0.4)	0 (0)	23.8 (+2.8)
Subtraction	34 (-23)	9 (0)	22.7 (+6.7)	0.4 (+0.4)	0 (0)	34 (+16)
Transfer Subtraction Combinations (Subtraction Items Not Practiced by Any Condition)						
Make-10	55.2 (-12.5)	16.1 (+7.2)	17.7 (+5.2)	0.5 (+0.5)	0.5 (+0.5)	9.9 (-1)
Near-doubles	54.3 (-9.1)	11.3 (-2.7)	22 (+8)	1.1 (0)	0 (0)	11.3 (+3.8)
Subtraction	44.8 (-23.4)	13 (+2.6)	25.5 (+10.9)	1 (+1)	0 (0)	15.6 (+8.8)

Hypothesis 3: The Efficacy of the Subtraction Training

F-Index. Planned comparisons between the structured subtraction group and the active control (structured make-10 and structured near-doubles) groups indicated significant differences in performance on both practiced, $F(1, 88) = 20.15, p < .001, g = 0.51$, and transfer items, $F(1, 89) = 4.72, p = .016, g = 0.32$. The effect sizes for practiced items and transfer items exceed the IES (2014) $g = 0.25$ criteria for substantively important practice favoring the subtraction intervention.

Fluency Rate. There were significant differences in performance between the structured subtraction group and the active control groups on the practiced, $F(1, 89) = 11.33, p < .001, g = 0.47$, and more importantly, the transfer items, $F(1, 89) = 2.80, p = .049, g = 0.25$. The effect sizes for practiced items and transfer items exceed the IES (2014) $g = 0.25$ criteria for substantively important practice favoring the subtraction intervention.

Hypothesis 4: Developmental Prerequisite Knowledge

Students who were not fluent on prerequisite plus-10 items were not fluent on the corresponding add-with-8 or -9 items or on the prerequisite doubles item were not fluent on the corresponding near-doubles items for both practiced and transfer items on the posttest. Results regarding subtraction items were inconclusive due to the low level of fluency on the developmental prerequisite addition complement items. See Table 21 for results.

Table 21

Relation among Performance on Hypothesized Developmental Prerequisites and Mental-Addition Posttest Fluency of Make-10 items (Frame A), Near-Doubles (Frame B) and Subtraction items (Frame C) by Condition: Make-10 condition vs. Near-Doubles [brackets] vs. Subtraction (parentheses).

Developmental Prerequisite	Mental-addition posttest item			
Frame A: Plus-ten items x adding-with-8 or -9 items				
	5 + 9		9 + 5	
	Not	Fluent	Not	Fluent
4 + 10 and 10 + 4	20 [19] (15)	8 [5] (4)	14 [15] (10)	14 [9] (9)
Fluent	3 [6] (13)	1 [1] (0)	2 [6] (12)	2 [1] (1)
Not	$p < .001$ [$p < .001$] ($p < .001$)		$p < .001$ [$p = .001$] ($p = .006$)	
	6 + 8 ¹		8 + 6 ¹	
	Not	Fluent	Not	Fluent
4 + 10 and 10 + 4	26 [22] (16)	2 [2] (3)	28 [21] (15)	0 [3] (4)
Fluent	4 [7] (13)	0 [0] (0)	4 [7] (13)	0 [0] (0)
Not	$p < .001$ [$p < .001$] ($p < .001$)		$p < .001$ [$p < .001$] ($p < .001$)	
	8 + 9 ¹		9 + 8 ¹	
	Not	Fluent	Not	Fluent
7+10 and 10+7	20 [20] (14)	5 [4] (5)	18 [17] (13)	7 [7] (6)
Fluent	6 [6] (12)	1 [1] (1)	6 [5] (13)	0 [2] (0)
Not	$p < .001$ [$p < .001$] ($p < .001$)		$p < .001$ [$p < .001$] ($p < .001$)	
Frame B: Doubles items x near-doubles				
	7 + 8		8 + 7	
	Not	Fluent	Not	Fluent
7 + 7	12 [16] (20)	7 [8] (3)	8 [13] (17)	11 [11] (6)
Fluent	11 [7] (9)	2 [0] (0)	10 [6] (9)	3 [1] (0)
Not	$p = .006$ [$p < .001$] ($p < .001$)		$p=.055$ [$p=.001$] ($p<.001$)	
	8 + 9 ¹		9 + 8 ¹	
	Not	Fluent	Not	Fluent
8 + 8	12 [16] (8)	6 [5] (3)	22 [31]	12 [1]
Fluent	14 [10] (18)	0 [0] (3)	1 [5]	1 [1]
Not	$p < .001$ [$p < .001$] ($p = .113$)		$p = .003$ [$p = .001$] ($p = .133$)	
Frame C: Addition complements x Subtraction items				
	11 - 7			
	Not	Fluent		
4 + 7	8 [3] (4)	0 [1] (3)		
Fluent	24 [25] (23)	0 [2] (2)		
Not	$p = .004$ [$p = .500$] ($p = .344$)			
	12 - 7 ¹			
	Not	Fluent		
5 + 7	3 [4] (5)	2 [3] (3)		
Fluent	26 [24] (23)	1 [0] (1)		
Not	$p=.313$ [$p=.063$] ($p=.109$)			

Table 21 continued

		12 – 5 ¹	
		Not	Fluent
7 + 5			
Fluent		3 [3] (6)	1 [1] (0)
Not		26 [23] (23)	2 [4] (3)
		$p = .500$ [$p = .500$] ($p = .254$)	
		14 – 7	
		Not	Fluent
7 + 7			
Fluent		8 [13] (4)	11 [11] (19)
Not		13 [5] (8)	0 [2] (1)
		$p = .004$ [$p = .004$] ($p = .188$)	
		13 – 4	
		Not	Fluent
9 + 4			
Fluent		7 [9] (7)	2 [0] (4)
Not		22 [22] (20)	1 [0] (1)
		$p = .035$ [$p = .002$] ($p = .035$)	

Note 1. "Fluent" for the developmental prerequisites indicates that a child met the criteria for fluency on one or both addition items; the "Not" indicates that a child did not meet fluency criteria on one or both prerequisites. A significant p value (McNemar test) confirms that a hypothesized prerequisite is a necessary condition for fluency with it.

Note 2. Groups appearing in **bold** practiced the prerequisite items in a given Frame.

Note 3. ¹ Indicates a transfer item.

Discussion

Implications

H1: Efficacy of Make-10 Training. H1 was partially supported. The adjusted mean F-Index score for make-10 students on practiced items improved +1.26 points representing responses indicative of slow reasoning strategies or deliberate undetermined strategies. The active controls both increased their performance but they remained in the category of responses indicative of slow undetermined strategies and counting. The percentage decrease of make-10 students offering slow, counted, or incorrect responses on the posttest was 21.3% compared to 4.9% and 15.6% for the near-doubles and subtraction groups, respectively; whereas the percentage of make-10 students using automatic or non-conscious reasoning strategies increased by 19.2% compared to 2.6% and 4.7% for the near-doubles and subtraction groups, respectively.

On transfer items, all groups remained in the category of correct, but slow or counted responses. The near-doubles students actually had a slight regression from their pretest performance on these items (-0.09 on F-Index and -0.01 on fluency rate). The percentage decrease of make-10 students offering slow, counted, or incorrect responses decreased by 8.1% compared to no change and 9.4% for the near-doubles and subtraction groups, respectively; while the percentage of make-10 students using automatic or subconscious reasoning strategies increased by 8.1% compared to 0.7% and 5.6% for the near-doubles and subtraction groups, respectively. Less than one-fifth of students in any training achieved Phase 3 on the posttest. Although the make-10 group did not improve significantly more than their peers who received regular classroom instruction, the make-10 group was efficacious as measured by effect size on both the practiced and the transfer items. Statistical significance is but one way of gauging efficacy.

H2: Efficacy of near-doubles training. The results for H2 were somewhat expected. Given that instruction on near-doubles items begins in grade 1, it could be rightly argued that obtaining statistical differences would be difficult due to a practice (i.e., classwork and homework) effect. The adjusted mean F-Index score for near-doubles students on practiced items improved +0.56 points remaining in the category of responses indicative of slow reasoning strategies or deliberate undetermined strategies. The active controls both increased their performance also offering the same category of responses as the near-doubles group. The percentage decrease of near-doubles students offering very slow, counted, or incorrect responses on the posttest was 15.8% compared to 7.7% and 16.7% for the make-10 and subtraction groups, respectively; while the percentage of near-doubles students using automatic or subconscious reasoning strategies increased by 11.6% compared to -0.2% and 5.7% for the make-10 and

subtraction groups, respectively. Unfortunately, the guided instruction was not more effective (marginally significant differences on both the F-Index and fluency rate measures) than regular classroom instruction; however, the Hedges' g statistics were above the substantively important 0.25 threshold favoring the near-doubles intervention for both the F-Index and fluency rate measures.

On transfer items, all groups remained in the category of correct, but slow or counted responses. The make-10 students actually had a slight regression from their pretest performance on these items (-0.27 on F-Index and -0.08 on fluency rate). The percentage decrease of near-doubles students offering slow, counted, or incorrect responses decreased by 12.9% compared to 5.2% and 7.3% for the make-10 and subtraction groups, respectively; while the percentage of make-10 students using automatic or subconscious reasoning strategies increased by 0.6% compared to -9.3% and 5.2% for the make-10 and subtraction groups, respectively. Less than one-fourth of students in any training achieved Phase 3 on the posttest. Although the near-doubles group did not improve significantly (marginally significant differences on practiced items) more than their peers who received regular classroom instruction, the near-doubles group was efficacious as measured by effect size on both the practiced and the transfer items. Statistical significance is but one way of gauging efficacy.

H3: Efficacy of Subtraction Training. H3 was supported. The adjusted mean F-Index score for subtraction students on practiced items improved +0.93 points representing responses indicative of slow reasoning strategies or deliberate undetermined strategies. The active controls both increased their performance but they remained in the category of responses indicative of slow undetermined strategies and counting. The percentage decrease of subtraction students offering slow, counted, or incorrect responses on the posttest was 23% compared to 10.2% and

5.2% for the make-10 and near-doubles groups, respectively; each group had identical gains of 0.4% increases of deliberate reasoning strategy usage; while the percentage of subtraction students using automatic or subconscious reasoning strategies increased by 16% compared to 4.3% and 2.8% for the make-10 and near-doubles groups, respectively. The guided instruction was more effective than regular classroom instruction for both the F-Index and fluency rate measures as gauged by both statistical significance and effect size.

On transfer items, all groups remained in the category of correct, but slow or counted responses. The percentage decrease of subtraction students offering slow, counted, or incorrect responses decreased by 20.8% compared to 5.3% and 11.8% for the make-10 and near-doubles groups, respectively; a slight increase of 1% in deliberate reasoning strategy usage compared to 0.5% and no change for the make-10 and near-doubles groups, respectively; while the percentage of subtraction students using automatic or subconscious reasoning strategies increased by 8.8% compared to -1% and 3.8% for the make-10 and near-doubles groups, respectively. The guided instruction was more effective in promoting transfer than regular classroom instruction for both the F-Index and fluency rate measures as gauged by both statistical significance and effect size.

H4: Developmental Prerequisites. H4 was supported for adding-with-8 or -9 items and near-doubles items, but inconclusive regarding subtraction items. Knowing the developmental prerequisites is instrumental in having success solving make-10 and near-doubles targeted items. With the exception of $8+9$ and $9+8$ for students studying subtraction-as-addition, generally, participants were not proficient on the targeted make-10 and near-doubles item without also being proficient on the prerequisite item(s). Consistent with the active control groups, students in the subtraction group who were not proficient on $9+4$ were not fluent on the related $13 - 4$. Except for $14 - 7$ most students were not fluent on the related addition developmental

prerequisites necessary to execute the subtraction-as-addition reasoning strategy leading to inconclusive results.

Qualifications and Limitations

Although the findings are promising, some limitations must be noted. (a) Generalizability is limited to the composition of the study samples namely: high achieving first graders, as well as third graders experiencing difficulties with mathematics. Thus, the results may not be representative of other categories of students. (b) The studies lacked a training condition that involved practice-only in a semi-random order. No group practiced all targeted items from all interventions; therefore, arguments cannot be made concerning the difference in structured discovery versus practice only. (c) A more complete assessment of each intervention's effectiveness could be ascertained if the samples spanned grades K-2. For grades K-1, the targeted material is presented for the first time, while in grade 2 the concepts receive the most instructional attention. Further, a multiple grade intervention may provide insights concerning the progression (or lack thereof) towards reaching the CCSSO (2011) goal of "by the end of Grade 2, know from memory all sums of two one-digit numbers." (d) There was a lack of an explanation for why an answer was correct or incorrect for each trial. For beginning learners, knowing why an answer does or does not make sense could inform their ability to rationalize why a particular reasoning strategy is effective. (e) The sample size per intervention was small which limits power. (f) Ideally, randomization should have occurred at the classroom level for each study in order to negate the possibility of an effect due to differential instruction. (g) Given that the studies focused on efficacy and were administered in a relatively structured research environment, standard classroom implementation is needed to gauge the true effectiveness of these interventions.

Conclusion

The Role of Guided Instruction. Consistent with the recommendations of NMAP (2008) and the number sense view: Structured discovery/practice can be an effective educational tool in promoting the learning of mathematical regularities and combination fluency especially as it relates to computer-assisted instruction (Clements & Sarama, 2012). Also, consistent with the NRC's recommendation (Kilpatrick et al., 2001), the learning of reasoning strategies can be furthered and often accelerated by conceptually based instruction. Each experiment had a significant result on a targeted reasoning strategy on practiced but more importantly unpracticed items.

Although the results regarding add-with-8 or -9 and near-doubles were not constant across the two experiments, significant F-Index and fluency rate measure differences indicate that purposeful practice on add-with-8 or -9 combinations are worthwhile experiences for advanced grade 1 students.

Regarding subtraction, the parts and the whole were labeled and color coded to draw attention to the complementary nature that addition-subtraction fact families shared the same parts and whole. The explicit labeling of common elements is omitted in some classroom curricula. The present subtraction intervention results suggest that the meaningful learning of a subtraction-as-addition strategy, which grade 3 students applied to unpracticed subtraction combinations, can greatly reduce the amount of practice and time needed to achieve fluency with basic subtraction combinations. Currently, there is no consensus as to what features should be stressed and highlighted to better assist students in their ability to meaningfully memorize their subtraction combinations. Such research is critical as educators attempt to make arithmetic equitable for all schoolchildren.

The Role of Practice Frequency. Practice as an instructional tool needs to be used purposefully and judiciously. The notion of hundreds or preferably thousands of repetitions as necessary to achieve (by rote) memorization as suggested by previous models and computer simulations of arithmetic learning (Shrager & Siegler, 1998; Siegler & Jenkins, 1989) are misguided. Purposeful instruction involves recognizing how addition and subtraction are interdependent operations (Piaget, 1964). When students are developmentally ready, instruction should involve integrating the related operations of addition and subtraction instead of treating arithmetic fact families as independent associations. Practicing combinations before a child has mastered the developmental prerequisites for a combination family may be ineffective in promoting meaningful learning.

Prior Knowledge and Salience. The complexity of the regularity (i.e., the difficulty of the steps involved) within a strategy may impact the learning of the targeted items. The ability to induce and assimilate the pattern or relation (i.e., salience) given a students' prior conceptual, factual, and procedural knowledge can be complicated by the level of fluency with the developmental prerequisite items. It could be argued that part of the difficulty at-risk primary age children have in learning the make-10 strategy stems from adding the appropriate amount to the larger addend (procedural) to form ten, decomposition of the smaller addend such that it can be easily added to the larger addend to form ten (factual), in certain cases using the associative property to rearrange the addends in working memory, recalling what remains from the smaller addend, and finally having to know the plus-ten facts (factual). Educators need to consider whether a child has mastered the developmental prerequisites for a particular combination, especially those involving subtraction, as well as the developmental prerequisites for mental arithmetic in general.

Future research efforts related to the learning of addition and subtraction in the early school years would greatly benefit from larger scale studies comparing practice-only versus guided instruction as well as guided interventions involving different themes (e.g., comparing use-ten to make-10 to practice only for add-with-8 or -9 items). Incorporating the multiple instruction model as measured by the F-Index for the purposes of gauging how close students are to achieving Phase 3 along with automaticity as measured by the fluency rate in a research environment, and later, in a standard classroom setting would go a long way towards enlightening educators concerning the obstacles that have yet to be overcome concerning the goal of achieving computational fluency on basic addition and subtraction items for all students before the end of Grade 3.

References

- Ackerman, P. T., Anhalt, J. M., & Dykman, R. A. (1986). Arithmetic automatization failure in children with attention and reading disorders: Associations and sequela. *Journal of Learning Disabilities, 19*, 222-232. doi:10.1177/002221948601900409
- Baroody, A.J. (1999). Children's relational knowledge of addition and subtraction. *Cognition and Instruction, 17*, 137-175. doi:10.1207/S1532690XCI170201
- Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences, 6*, 1-36. doi:10.1016/1041-6080(94)90013-2
- Baroody, A. J. (1985). Mastery of the basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education, 16*, 83–98. doi:10.2307/748366
- Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review, 3*, 225-230. doi:10.1016/0273-2297(83)90031-X
- Baroody, A. J., & Purpura, D. J. (in preparation). Number and operations. In J. Cai (Ed.), *Mathematics education and research handbook*. Reston, VA: National Council of Teachers of Mathematics.
- Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Review, 15*, 69-79 (Special issue on "Pathways to Mathematical Learning Disabilities" guest edited by M. Mazzocco). doi:10.1002/ddrr.45
- Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development, 20*, 80-128. doi:10.1080/10409280802206619
- Baroody, A. J., Torbeyns, J., & Verschaffel, L. (2009). Young children's understanding and

application of subtraction-related principles. *Mathematical Thinking and Learning*, 11, 2-9.

doi:10.1080/10986060802583873

Baroody, A. J., Thompson, B., & Eiland, M. (2008). *Fostering the fact fluency of grade 1 at-risk children*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Baroody, A. J., Eiland, M. D., Paliwal, V., Priya-Bajwa, N., & Baroody, S. C. (2010). *Fostering at-risk primary-grade children's fluency with basic addition combinations*. Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

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. doi:10.3102/0002831212473349

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.
doi:10.1080/07370008.2012.720152

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300. doi:10.2307/2346101

Brownell, W. A. (1935). Psychological considerations in the learning and the teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic* (Tenth yearbook, National Council of Teachers of Mathematics, pp. 1–31). New York: Bureau of Publications, Teachers College, Columbia University.

- Brownell, W. A. (1941). *Arithmetic in grades I and II: A critical summary of new and previously reported research*. Durham, NC: Duke University Press.
- Buckingham, B. R. (1927). Teaching addition and subtraction facts together or separately. *Educational Research Bulletin*, 6, 228-229, 240-242.
- Bullock, J. K. (1971-present). *Touch Math*. Colorado Springs, CO: Innovative Learning Concepts, Inc.
- 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. doi:10.1037/0012-1649.34.5.882
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, 179–202. doi:10.2307/748348
- Clements, D. H., & Sarama, J. (2012). Learning and teaching early and elementary mathematics. In Carlson, J., & Levin, J. (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. doi:10.1016/j.cogdev.2009.09.005

- Dowker, A. (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. 243-265). Mahwah, NJ: Erlbaum.
- Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*, 23, 45-55. doi: 10.2307/749163
- Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction. *Cognition*, 123, 392–403. doi:10.1016/j.cognition.2012.02.008
- Fuson, K. C. (2006). *Math expressions*. Boston, MA: Houghton Mifflin.
- Geary, D. C. (1996). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association. (Original work published 1994). doi:10.1037/10163-000
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, 33, 18–28. doi:10.1177/002246699903300102
- Goldman, S. R., Pellegrino, J., & Mertz, D. L. (1988). Extended practice of basic addition facts: Strategy changes in learning disabled students. *Cognition and Instruction*, 5, 223–265. doi:10.1207/s1532690xci0503_2
- Hatano, G. (2003). Forward. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. xi–xiii). Mahwah, NJ: Erlbaum.
- Heavey, L. (2003). Arithmetical savants. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 409–434). Mahwah, NJ: Erlbaum.

- 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. doi:10.2307/30034895
- Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Erlbaum.
- Institute of Education Sciences (2014). *What Works Clearinghouse: Procedures and standards handbook (Version 3.0)*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf
- James, W. (1958). *Talks to teachers on psychology and to students on some of life's ideals*. New York, NY: W. W. Norton & Company. (Talk originally given in 1892.)
- Jerman, M. (1970). Some strategies for solving simple multiplication combinations. *Journal for Research in Mathematics Education*, 1, 95–128. doi:10.2307/748856
- Jordan, N. C. (2007). The need for number sense. *Educational Leadership*, 65(2), 63–66.
- Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, 85, 103–119. doi:10.1016/S0022-0965(03)00032-8
- Jordan, N. C., Hanich, L. B., & Uberti, H. Z. (2003). Mathematical thinking and learning difficulties. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts*

- and skills: Constructing adaptive expertise* (pp. 359–383). Mahwah, NJ: Erlbaum.
- Katona, G. (1967). *Organizing and memorizing: Studies in the psychology of learning and teaching*. New York, NY: Hafner. (Work originally published in 1940.)
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, D.C.: National Academy Press.
- Kraner, R. E. (1980). Math deficits of learning disabled first graders with mathematics as a primary and secondary disorder. *Focus on Learning Problems in Mathematics*, 2(3), 7–27.
- Larson, N. (2008). *Saxon math 1*. Austin, TX: Harcourt Achieve
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K. Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (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.
- Mason, J., & Spence, M. (1999). Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics*, 38, 135–161.
doi:10.1023/A:1003622804002
- Mead, C. D., & Sears, I. (1916). Additive subtraction and multiplicative division tested. *Journal of Educational Psychology*, 7, 261-270. doi:10.1037/h0075339
- Muarata, 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.
doi:10.1207/s1532690xci2202_2
- 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, D.C.: U.S. Department of Education.
- National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Washington, D.C.: National Academy Press.
- National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.
- Olander, H. T. (1931). Transfer of learning in simple addition and subtraction. *Elementary School Journal*, 31, 427–437. doi:10.1086/456594
- Perry, M., VanderStoep, S. W., & Yu, S. L. (1993). Asking questions in first-grade mathematics classes: Potential influences on mathematical thought. *Journal of Educational Psychology*, 85, 31-40. doi:10.1037/0022-0663.85.1.31
- Perry, M., McConney, M., Flevares, L. M., Mingle, L. A., & Hamm, J. V. (2011). Engaging first graders to participate a students of mathematics. *Theory Into Practice*, 50, 293-299. doi:10.1080/00405841.2011.607388
- Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered* (pp. 7–20). Ithaca, NY: Cornell University.
- Purpura, D. J., Baroody, A. J., Eiland, M. D., & Reid, E. E. (2014). Fostering reasoning strategies with the most basic sums via computer-assisted instruction: The value of guided training varies with combination family. Manuscript submitted for publication.
- Rathmell, E. C. (1978). Using thinking strategies to teach basic facts. In M. N. Suydam & R. E. Reys (Eds.), *Developing computational skills* (1978 Yearbook, pp. 13–50). Reston, VA: National Council of Teachers of Mathematics.

- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93, 346–362. doi:10.1037/0022-0663.93.2.346
- Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction*, 1, 217–244. doi:10.1207/s1532690xci0102_3
- Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.
- Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, 9, 405–410. doi:10.1111/1467-9280.00076
- Siegler, R.S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- Skemp, R. R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher*, 26(3), 9–15.
- Skemp, R. R. (1979). *Intelligence, learning, and action*. Chichester, UK: Wiley.
- Skemp, R. R. (1987). *The psychology of learning mathematics*. Hillsdale, NJ: Erlbaum.
- Smith, J. H. (1921). Arithmetic combinations. *The Elementary School Journal*, 10, 762–770.
- Steinberg, R. M. (1985). Instruction on derived facts strategies in addition and subtraction. *Journal for Research in Mathematics Education*, 16, 337–355. doi:10.2307/749356
- Suydam, M., & Weaver, J. F. (1975). Research on mathematics learning. In J.N., Payne (Ed.), *Mathematics Learning in Early Childhood* (pp. 44–67). Reston, VA: NCTM.
- Suzzalo, H. (1911). *The teaching of primary arithmetic: A critical study of recent tendencies*

method. Boston: Houghton Mifflin Company.

Swenson, E. J. (1949). Organization and generalization as factors in learning, transfer and retroactive inhibition. In *Learning theory in school situations* (No. 2). Minneapolis: University of Minnesota Press.

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

Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.

Thorndike, E. L. (1918). *Educational psychology: Briefer course*. New York: Teachers College Columbia University.

Threlfall, J., Frobisher, L., & MacNamara, A. (1995). Inferring the use of recall in simple addition. *British Journal of Educational Psychology*, 65, 425-439. doi:10.1111/j.2044-8279.1995.tb01163.x

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. doi:10.1207/s1532690xci2301_1

University of Chicago School Mathematics Project. (2005). *Everyday mathematics teacher's lesson guide (Volume 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.

Wertheimer, M. (1959). *Productive thinking*. New York, NY: Harper & Row. (Original work published 1945.)

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604.

doi:10.1037/0003-066X.54.8.594

Woods, S. S., Resnick, L. B., & Groen, G. J. (1975). An experimental test of five process models for subtraction. *Journal of Educational Psychology*, 67, 17–21. doi:10.1037/h0078666

Appendix A: Curriculum Details (Baroody, Purpura, Eiland, & Reid, 2014)

A Comparison of How Seven Curricula Teach Addition and Subtraction

Curriculum	Feature					
	Explicit complement principle ¹	Explicit subtract-as-add strategy ¹	Same numbers or fact families ¹	Empirical inversion ¹	Explicit part-whole relations ¹	Known part in the same position ²
<i>Bridges in Mathematics</i> (Math Learning Center, 2009)	-	-	-	✓✓ ^a	-	N/A
<i>Everyday Mathematics</i> (UCSMP, 2005)	✓	✓✓ ^c	✓✓ ^d	-	✓	-
<i>Go Math 2012</i> (Houghton Mifflin Harcourt)	✓✓	✓✓	✓✓	✓	✓	✓
<i>Math Connects</i> (Macmillan/McGraw Hill, 2009)	✓✓ ^f	✓✓ ^f	✓✓ ^f	- ^g	✓	✓
<i>Math Expressions</i> (Fuson, 2006)	-	✓	✓✓	- ^h	✓	-
<i>Saxon Math</i> (Larson, 2008)	-	✓ ⁱ	✓✓ ^j	✓ ^j	-	N/A
<i>Scott Foresman Mathematics</i> (Charles & Hall, 2008)	✓✓ ^k	- ^k	✓✓ ^k	- ^l	✓	N/A

¹ A double check (✓✓) = clearly a prevalent characteristic and explicit¹; a single check (✓) = an infrequent or implicit characteristic; and a dash (-) = not characteristic.

² A double check (✓✓) = common part in the same position for related addition and subtraction equations; a single check (✓) = sometimes in the same position, sometimes not; and a dash (-) = principle/strategy always illustrated with common part in a different position. N/A = subtraction-as-addition strategy not explicitly taught.

^aFor example, both operations are introduced as “hopping along a number line.” An addition item such as 5+3 is represented as hopping to the right 5 and then 3; 5+3=8 is followed by 8-3, which is represented as hopping from 8 in the opposite direction (to the left) three times (8-3=5).

^bLesson 4.11 contains the following general hint (p. 309): “Some facts lead to other facts. There are not really that many different facts to learn or memorize. You will learn some easy ‘shortcuts’ in later lessons.” Lesson 5.10, encourages the if-then shortcut with commuted combinations (p. 379): “If you know one fact, then you also know its turn-around fact. Although implied, an if-then shortcut (the complement principle) is not explicitly specified for subtraction.

^cThe subtraction-as-addition strategy is explicitly recommended in at least one subtraction lessons. For Lesson 6.5 (Using the Addition/Subtraction Facts Table to Solve Subtraction Problems), the instruction specify (p. 510): “To find the answer to 15 - 9, ask yourself: ‘9 plus what number is 15?’” However, an analogous strategy is not explicitly recommended when working with other models (fact triples such as 3, 5, 8 represented by dominoes or fact triangles). A teacher note In Unit 6, specifies (p. 500): “For many first graders, it is helpful to think about 8 - 5 = ? as 5 + what number? = 8. This approach encourages ‘adding up’ to subtract, a strategy that also works well with multidigit numbers.” Encouraging thinking of subtraction as addition and using a counting-up strategy to solve for the missing addends can be a meaningful and useful for step toward, but not the same as recommending, using known sums as a vehicle for shortcutting subtraction computation—for deducing differences.

Appendix A continued

^d Unit 6.3 introduces addition/subtraction fact families (four facts related by the complement principle such as $3+5=8$, $5+3=8$, $8-3=5$, and $8-5=3$) and finding both sums and differences using dominoes (fact triples: 3, 5, 8) and looking for the equivalent names for sums and differences using the same Addition/Subtraction Fact Table.

Unit 6.4 does the same sums using fact triangles. Unit 6.5 uses Addition/Subtraction Fact Table to determine differences and Fact Triangles to generate addition-subtraction fact families.

^e For Unit 3, teachers are instructed: "Point out that the domino has a part with 3 dots and a part with 5 dots and that the whole domino has 8 dots" (p. 234), and children practice translating various dominoes into part-part-whole diagrams. However, this is not continued later (e.g., Unit 6.2) when the focus is on addition facts and never related to subtraction facts. Indeed, there is no mention of "parts" and "wholes" in Unit 6.2. In brief, the models used in this unit, such as Fact Triangles only implicitly represent part-part-whole relations of related addition and subtraction combinations.

^f Exercises include, for instance, (a) "Think $5+9=\square$ so $14-9=\square$ "; (b) Circle the addition fact that will help you subtract $12-9=\square$ $5+7=12$ or $5+6=11$, and (c) "To find $16-7$, I think $\square+7=16$? Related addition and subtraction facts are defined as having the same numbers (e.g., $1+6=7$ and $7-6=1$ or $3+7=10$ and $10-3=7$).

^g Exercises include showing two collections (e.g., 6 bees in a circle and 3 bees in a line) and asking a child to write one addition sentence ($6+3=9$ or $3+6=9$) and one subtraction sentence ($9-3$ or $9-6=9$). This might implicitly involve empirical inversion if a child thought: "6 and 3 more is 9, and 9 take away the 3 is, oh, 6 again."

^h For Unit 2, a teacher is encouraged to show that $8-3=5$ and an iconic representation (ooo|oooo) are the "reverse story of $5+3=8$. However, this relation is not shown in the textbook or a (seatwork or homework) worksheet, with one exception (an exercise for students "on [grade] level" in the "Extending the Lesson—Differentiated Instruction/Activities for Individualizing" section).

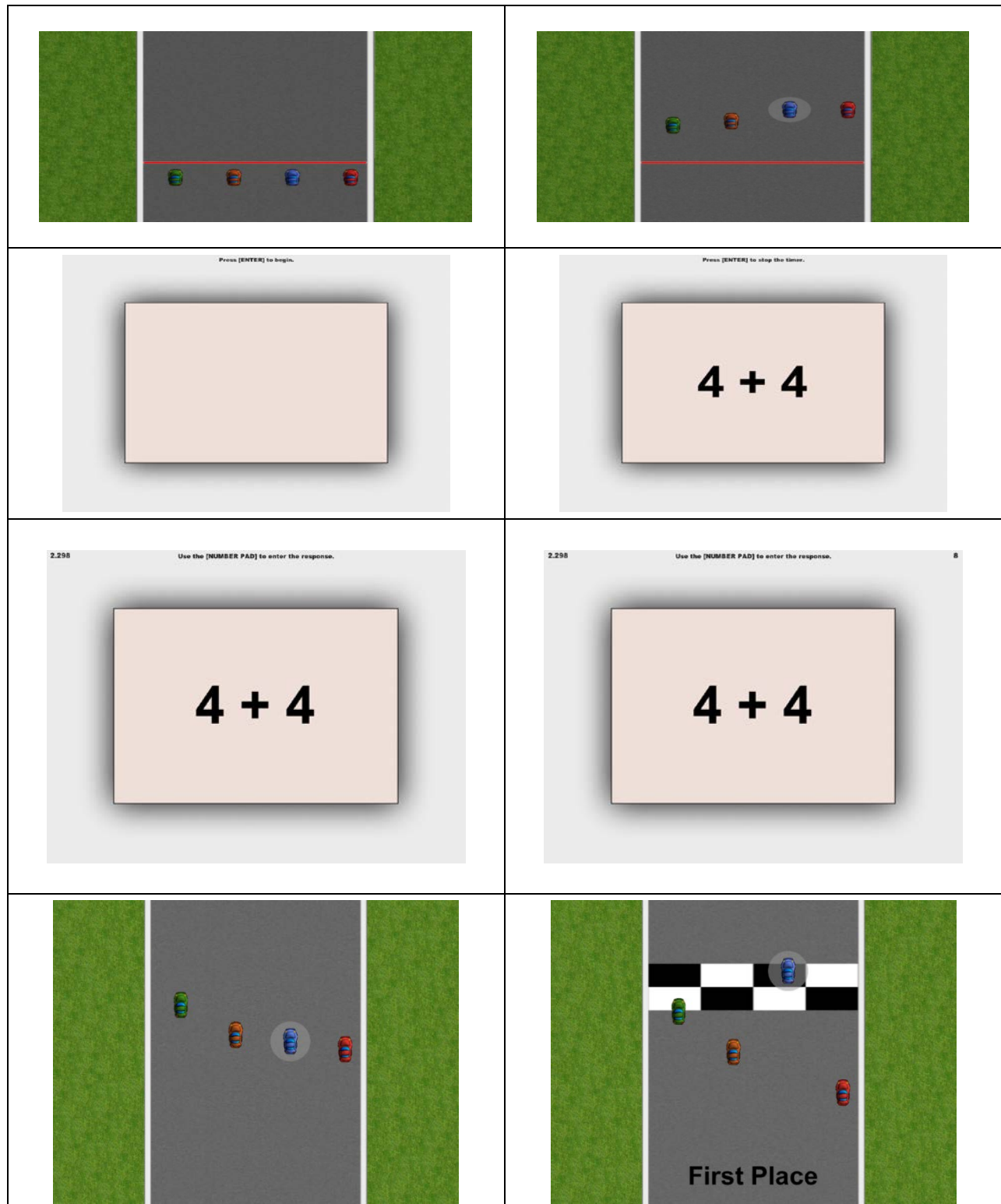
ⁱ The program inconsistently related subtraction to addition. On one hand, for example, although adding with 1 is practiced immediately before subtracting by 1 is introduced in Lesson 44, the two are not related. Similarly, the addition facts with sums to 10 are introduced in Lessons 94 and 95. Although Lessons 101 and 102 begin with practicing these sums, they are not related to the focus of the lesson—subtracting a number from 10. Although Lesson 121, which introduces the "Differences of 1 Subtraction Facts, relates $10-9=1$ to using addition to check a difference ($1+9=10$), add-with-1 combinations are not recommended as vehicle for determining the differences of such subtraction combinations. In Lesson 68 adding 2 to an odd or an even number with linking cubes is reviewed immediately before subtracting 2 from odd and even numbers is introduced with cubes. The former is summarized by the rule that "adding 2 is like saying the next add even or odd number" and the later is summarized as "subtracting 2 is like saying the even or odd numbers backwards by 2's." Although these rules may be useful, addition again is used as a shortcut for determining differences. On the other hand, in Lesson 129, unknown subtraction combinations are linked to known addition combinations. For "subtracting half of a double" such as $14-7$, $12-6$, and $8-4$, children are asked: "What do you notice about each these problems?" [they are the doubles going the other way]...How can we remember these answers? We will call these problems the 'subtracting half of a double facts.' See Footnote i for additional examples.

^j In Lesson 132, blue and red linking cubes are used to model $4+1=5$ and then $5-1=4$, $1+4=5$ and then $5-4=1$ (examples of empirical inversion) to introduce the concept of "addition and subtraction families" and as a method for learning $9-4$, $9-5$, $9-3$, and $9-6$. Fact families, but not empirical inversion, are used to introduce the $7-3$, $7-4$, $8-3$, and $8-5$ subtraction facts in Lesson 134.

^k Children are told: "You can use addition to help you subtract" and are encouraged to "think if $5+8=13$, then $13-8$ " and "think $5+8=?$, so $13-8=?$ (the complement principle). Although text points out that "the sum of an addition fact is the first number in [a related] subtraction fact" (p. 138), nowhere is it explained that the addend in addition the equation that does not appear in the related subtraction equation is the difference or answer for the latter. Put differently, instruction does not suggest how the complement principle can be used to determine unknown differences by using known addition combinations (e.g., solve $13-8=?$ By thinking "what do I have to add to 8 to make 8?").

^l One enrichment worksheet illustrates, for example, 1 green interlocking cube added to 4 white cubes with equation $4+\square=5$ and the statement, "Add the green cube. 4 and 1 more is 5." Immediately below this is the equation $5-\square=4$ and the statement, "Take away the green cube. 5 take away 1 more is 4." Unfortunately, as an enrichment activity, it may not be used in most classes.

Appendix B: Mental Arithmetic Test Screen Shots



Appendix C: Procedures for Determining the Use of a Response Bias

Response bias determination were done separately for each session and based on all 20 items given in a session. This was done because it is not uncommon for children to change a predominant strategy in a later session (e.g., adopt a more advanced strategy). The determination was done for the two sets presented in a session instead of by one set to insure a sufficient number of examples of a response bias to make a reliable determination.

Table 1 lists all the response biases that were actually found. In addition to these strategies, scorers looked for other common response bias such as *stating a favorite* (e.g., typically responding to various items with “five”) or “counting” (e.g., responding “one” to an item, “two” to the next item, “three” to the following item, and so forth).

1. If a single response bias was apparent, then a scorer scored the response to each item in one of five ways:
 - a. Possible inappropriate use of a strategy (resulting in an incorrect answer);
 - b. Possible appropriate use of a strategy (resulting in a correct answer);
 - c. Other incorrect estimates (error not attributable to the strategy);
 - d. Other correct estimates (correct response not attributable to the strategy); and
 - e. Use of concrete or abstract counting strategy or no response.

If two (or more) response biases were apparent, then the scorer scored the response to each item for each separately using the categories listed in guideline #3. The more prevalent strategy was used for further analyses.

2. In order to rise to the level of a response bias, four criteria had to be met:
 - a. Total inappropriate uses is half or more of all estimation errors: $a_{TOT} \div (a_{TOT} + c_{TOT}) \geq .50$;
 - b. Total inappropriate uses is more than 25% of all trials: $a_{TOT} \div 24 > .25$ (MINIMUM of 6 trials);
 - c. Total uses of strategy is half or more of all estimates: $(a_{TOT} + b_{TOT}) \div (a_{TOT} + b_{TOT} + c_{TOT} + d_{TOT}) \geq .50$; and
 - d. Total uses of strategy is half or more of all trials: $(a_{TOT} + b_{TOT}) \div 24 \geq .50$ (MINIMUM of 10 trials).

These criteria further ensured a consistency of strategy use over a variety of items and a sufficient number of trials to ensure a reliable determination of a response bias.

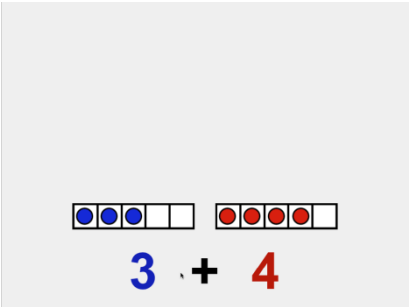
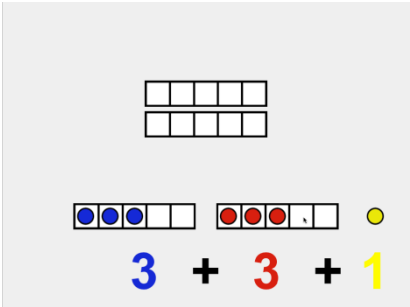
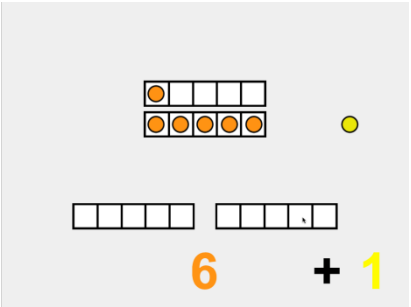
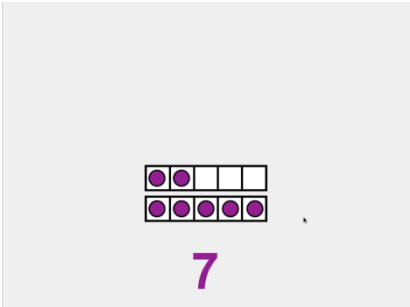
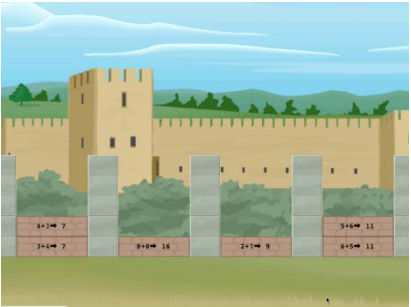
Appendix C continued: Response Bias (number of false positives) Identification by Type, Condition, and Testing Session

	Pretest		Posttest			
	Session 1 Sets 1 & 2	Session 2 Sets 3 & 4	Session 1 Sets 1 & 2	Session 2 Sets 3 & 4	Session 3 Sets 5 & 6	Session 4 Sets 7 & 8
Participant	Structured Make-10 Condition					
1	-	-	Number After Larger (3)	Number After Larger (1)	-	-
2	Number After (3)	-	-	-	-	-
3	-	-	-	-	Favorite Number (2)	-
4	Number After (3)	-	-	-	-	-
5	-	-	-	-	Favorite Number (2)	-
	Structured Near-Doubles Condition					
6	Number After Larger (3)	Number After Larger (1)	-	-	-	-

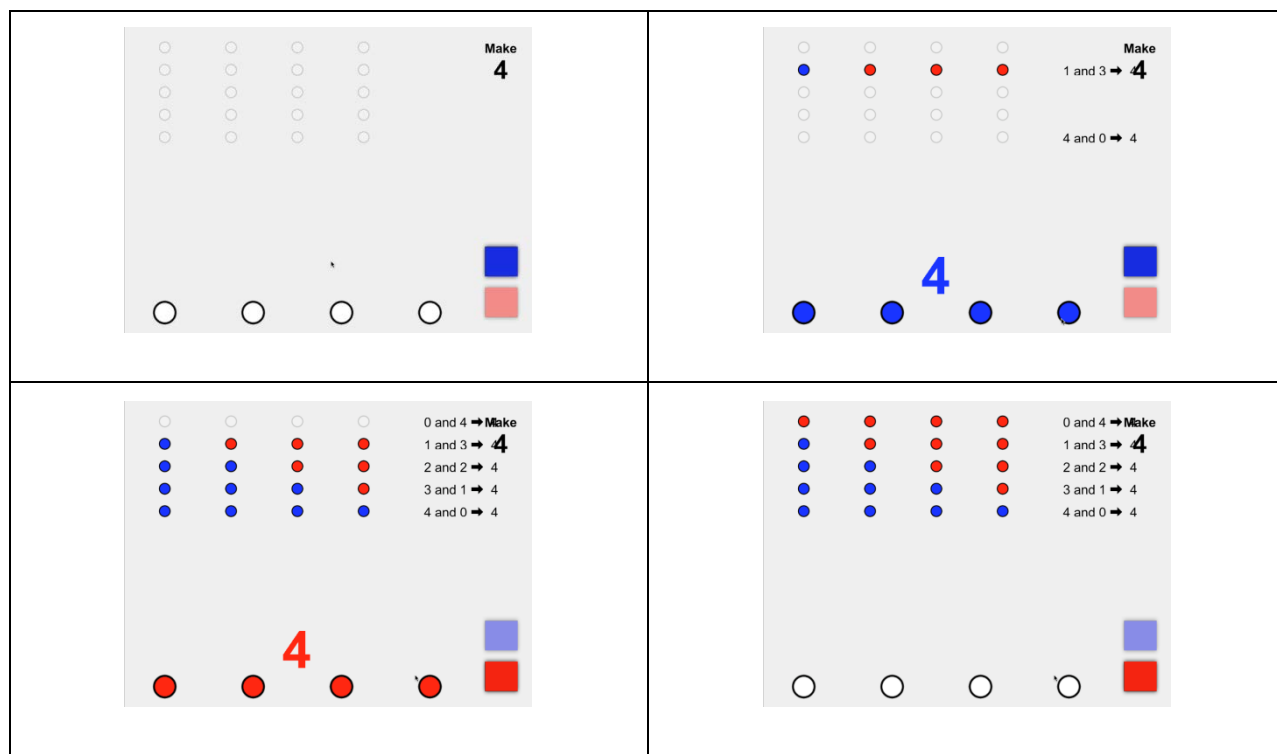
Appendix D: Mental Arithmetic Training Screen Shots

<div><p>Press [ENTER] to stop the timer.</p><div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>9</div><div>+</div><div>5</div></div></div></div></div>	<div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>9</div><div>+</div><div>5</div></div></div></div>
<div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>9</div><div>+</div><div>1</div><div>+</div><div>4</div></div></div></div>	<div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>10</div><div>+</div><div>4</div></div></div></div>
<div><p>Press [ENTER] to begin.</p><div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>14</div></div></div></div></div>	<div><div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div><div><div><div></div><div></div><div></div><div></div><div></div></div></div></div></div><div><div>14</div></div></div>

Appendix D continued

<p>1.89 Use the [NUMBER PAD] to enter the response.</p> $3 + 4$	
	
	

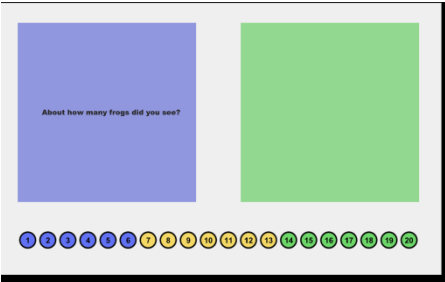
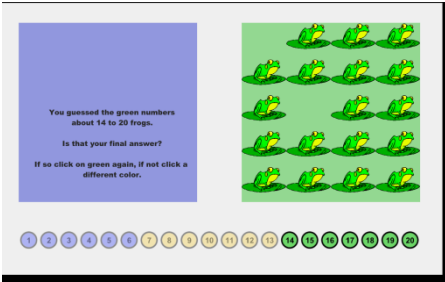
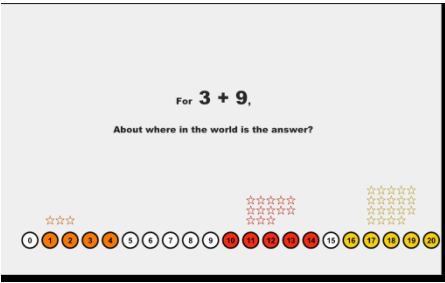

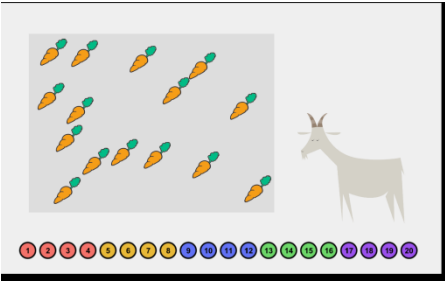
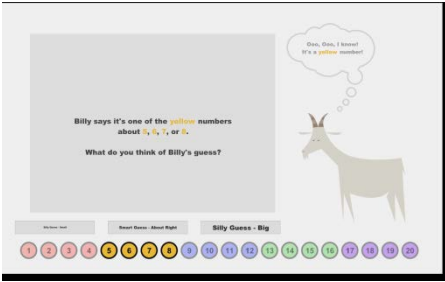
Appendix D continued



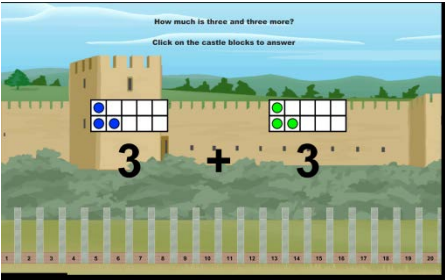
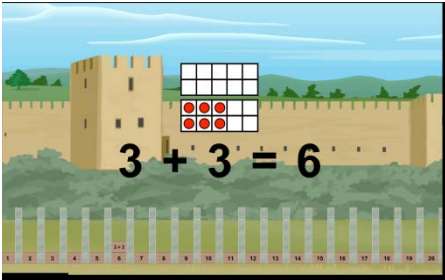
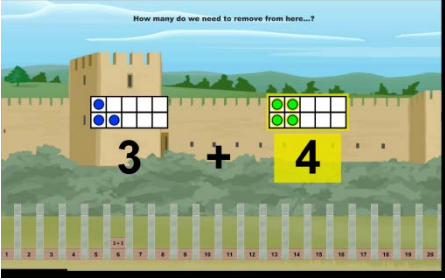
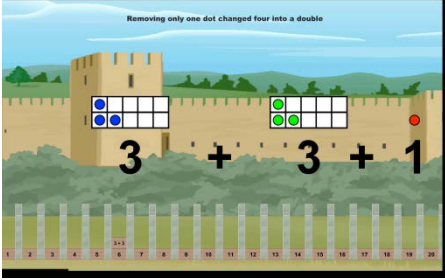
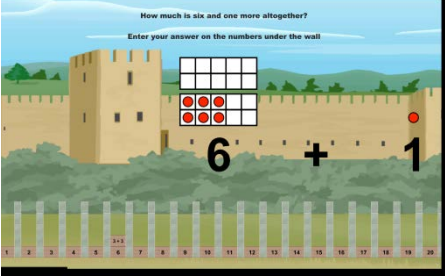
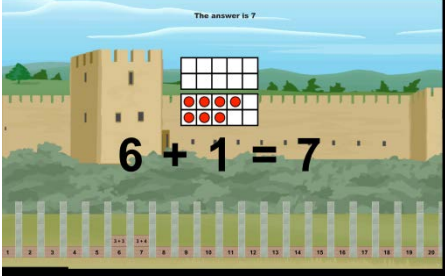
Appendix E: Summary of Mental Arithmetic Training and Practice Provided by Curricula Used in the Schools

Curriculum	Add-with-8 and -9 items	Near-doubles	Subtraction
<p><i>Everyday Mathematics</i> (UCSMP, 2005)</p> <p>(Note. Information cited in this table appears in Teacher's Manual but not student worksheets.)</p>	<p>In Lesson 2.4, students are taught to “add 10 instead of 9, and then count back 1. Add 10 instead of 8, and then count back 2.” This aligns with a <i>use-10</i> strategy.</p>	<p>In Lesson 2.5, students are taught to use a related doubles fact “for the fact $4 + 5$, think ‘$4 + 4 = 8$ and 1 more is 9.’ Or think ‘$5 + 5 = 10$ and 1 less is 9.’”</p>	<p>The subtraction-as-addition strategy is explicitly taught in Lesson 6.5 (Using the Addition/Subtraction Facts Table to Solve Subtraction Problems). The instructions specify (p. 510): “To find the answer to $15-9$, ask yourself: ‘9 plus what number is 15?’” However, an analogous strategy is not explicitly recommended when working with other models (fact triples such as 3, 5, 8 represented by dominoes or fact triangles). A teacher note for Unit 6, specifies (p. 500): “For many first graders, it is helpful to think about $8-5 = ?$ as $5 + \text{what number} = 8$. This approach encourages ‘adding/counting up’ to subtract,” (a meaningful step toward, but not the same as, a subtraction-as-addition strategy). Unit 6 introduces addition/subtraction fact families (four facts related by the complement principle such as $3+5=8$, $5+3=8$, $8-3=5$, and $8-5=3$) and finding both sums and differences using dominoes (fact triples: 3, 5, 8) and determining and looking for the equivalent names for sums and differences using the same Addition/Subtraction Fact Table and fact triangles.</p>
<p>Math Expressions (Fuson, 2006)</p> <p>(Note. Information cited in this table appears in the Teacher's Manual but not student worksheets.)</p>	<p>Unit 1 Lesson 6, begins with break-aparts using 10. In Lesson 7, Partner Houses with sums that total 2 to 10. In Lesson 8, Math Mountains are related to the floors of Partner Houses. In Lesson 15, a partner's fingers are used to model the number needed to complete 10 with the larger addend and with the other hand the amount over 10 in the teen number (e.g., for $8 + 6$; on the left hand as student produces 2 fingers to complete the 10 while on the right hand, the 4 fingers necessary to complete the 6. 4 fingers over 10 equals 14. The make-10 strategy is practiced extensively as daily warm-ups throughout Unit 2 with $7 + n$, $8 + n$, and $9 + n$ items.</p>	<p>Unit 1 Lesson 11, students find totals using the doubles plus or doubles minus 1 strategy. Offers the hint “Remind children that when they use the doubles plus 1 strategy, they only increase one of the addends in the doubles fact.” Practice with small addends (under 5) and larger addends (5 to 9).</p>	<p>In Unit 1 Lesson 14, students are encouraged to count on to solve addition, mystery addition, and subtraction problems. Math Mountains are used to model problem elements. Students are charged with finding the missing “part” (e.g., $5 + \square = 8 - \square = 3$) or the missing “total”. In Lesson 15, students make a ten to solve mystery and subtraction story problems. A partner's fingers are used to model the number needed to complete 10 with the larger addend and with the other hand the amount over 10 in the teen number. Lesson 16 relates addition and subtraction teen totals “I think $14 - 8 = 6$ or $8 + \text{how many} = 14$.” Several in-class and homework opportunities are provided to practice determining the missing value. Lesson 19 forms equations from Mountain Math. In equations, “squiggles” are used to denote parts; “T” is placed under the total.</p>

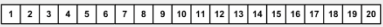
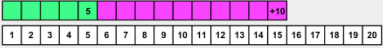
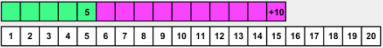
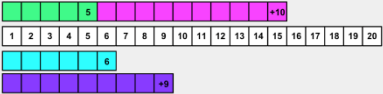
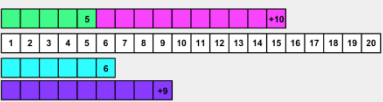
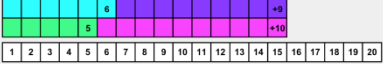
Appendix F: Experiment 1 Captions

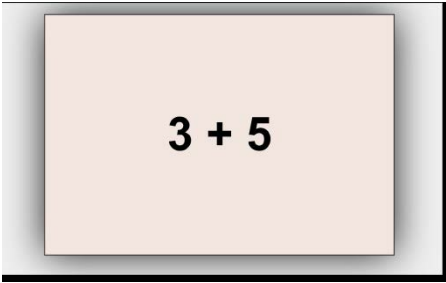
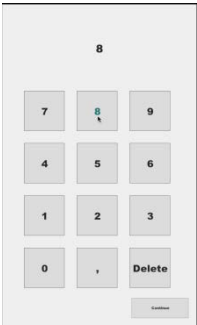
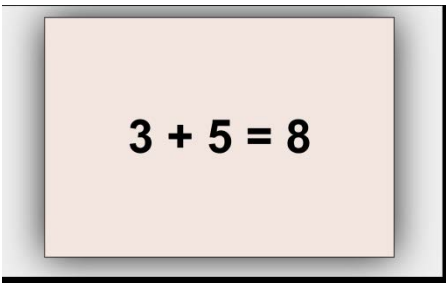


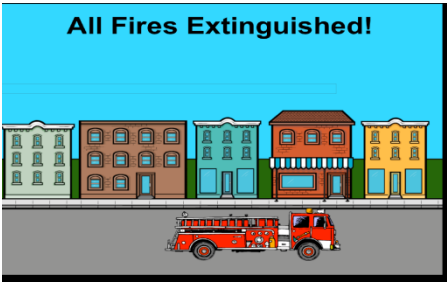
Appendix F continued

<p>How much is three and three more? Click on the castle blocks to answer</p>  <p>$3 + 3$</p>	 <p>$3 + 3 = 6$</p>
<p>How many do we need to remove from here...?</p>  <p>$3 + 4$</p>	<p>Removing only one dot changed four into a double</p>  <p>$3 + 3 + 1$</p>
<p>How much is six and one more altogether? Enter your answer on the numbers under the wall</p>  <p>$6 + 1$</p>	<p>The answer is 7</p>  <p>$6 + 1 = 7$</p>

Appendix F continued

<p>$5 + 10 = ?$</p> 	<p>five and ten more is fifteen</p> <p>$5 + 10 = 15$</p> 
<p>Does five and ten more is fifteen help you to answer...</p> <p>$5 + 10 = 15$</p>  <p>$6 + 9 = ?$</p>	<p>Does five and ten more is fifteen help you to answer what six and nine more is?</p> <p>No, because six and nine is NOT fifteen</p> <p>Yes, because six and nine is also fifteen</p>  <p>$6 + 9 = ?$</p>
<p>Let's see if five and ten more helps to answer six and nine more If six and nine more is also fifteen</p>  <p>$6 + 9 = ?$</p>	<p>So, yes, five and ten more is fifteen helps to answer what six and nine more is because six and nine more is also fifteen</p> <p>$5 + 10 = 15$</p>  <p>$6 + 9 = 15$</p>



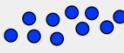


Appendix F continued

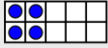

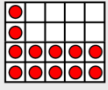
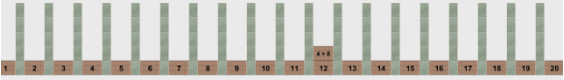

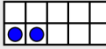



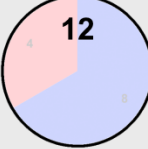
Appendix G: Experiment 1 Response Biases Identification by Type, Condition and Testing Session

	Pretest		Posttest	
	Session 1	Session 2	Session 1	Session 2
Participant	Structured Make-10 Condition			
1	-	-	Make-a-teen (4)	-
2	-	Make-a-teen (3)	-	-
3	-	Make-a-teen (3)	-	-
4	Make-a-teen (4)	-	-	-
5	Make-a-teen (4)	Make-a-teen (3)	-	-
6	-	-	-	Make-a-teen (3)
7	-	-	Make-a-teen (4)	Make-a-teen (3)
8	-	-	Make-a-teen (4)	-
	Structured Near-doubles Condition			
9	-	-	Make-a-teen (4)	Make-a-teen (3)
10	-	-	Make-a-teen (4)	-
11	Make-a-teen (4)	-	-	-
12	-	Make-a-teen (3)	-	-
13	Make-a-teen (4)	-	-	-
14	Make-a-teen (4)	-	-	-
15	-	Make-a-teen (3)	-	-

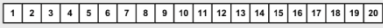
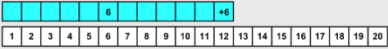
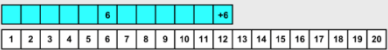
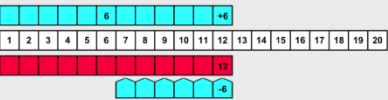

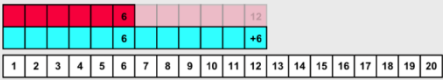
Appendix H: Experiment 2 Screen Captions

<p>Mya went first. She grabbed the ball and threw it at the target. She scored 5 points! Then, she took her second turn and scored 7 points. How many is 5 points and 7 more points altogether?</p>  <div style="display: flex; justify-content: space-around; margin-top: 20px;"> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">5+7</div> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">5x7</div> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">7+5</div> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">7x5</div> </div>	<p>You can solve this problem anyway you want. Using your head, using your fingers or using the numberline. On the numberline click on the first number then the second number. When you have your answer press the "Ready to Answer" button.</p> <div style="text-align: center; margin-top: 20px;"> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr> <td style="width: 20px; height: 20px;">0</td><td style="width: 20px; height: 20px;">1</td><td style="width: 20px; height: 20px;">2</td><td style="width: 20px; height: 20px;">3</td><td style="width: 20px; height: 20px;">4</td><td style="width: 20px; height: 20px;">5</td><td style="width: 20px; height: 20px;">6</td><td style="width: 20px; height: 20px;">7</td><td style="width: 20px; height: 20px;">8</td><td style="width: 20px; height: 20px;">9</td><td style="width: 20px; height: 20px;">10</td><td style="width: 20px; height: 20px;">11</td><td style="width: 20px; height: 20px;">12</td><td style="width: 20px; height: 20px;">13</td><td style="width: 20px; height: 20px;">14</td><td style="width: 20px; height: 20px;">15</td><td style="width: 20px; height: 20px;">16</td><td style="width: 20px; height: 20px;">17</td><td style="width: 20px; height: 20px;">18</td><td style="width: 20px; height: 20px;">19</td><td style="width: 20px; height: 20px;">20</td> </tr> </table> <div style="font-size: 48px; font-weight: bold; margin-top: 10px;">5 + 7</div> <div style="display: flex; justify-content: center; gap: 20px; margin-top: 20px;"> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">Ready To Answer</div> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0;">Start Over</div> </div> </div>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																		
<div style="text-align: center; margin-bottom: 10px;"> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr> <td style="width: 20px; height: 20px;">0</td><td style="width: 20px; height: 20px;">1</td><td style="width: 20px; height: 20px;">2</td><td style="width: 20px; height: 20px;">3</td><td style="width: 20px; height: 20px;">4</td><td style="width: 20px; height: 20px;">5</td><td style="width: 20px; height: 20px;">6</td><td style="width: 20px; height: 20px;">7</td><td style="width: 20px; height: 20px;">8</td><td style="width: 20px; height: 20px;">9</td><td style="width: 20px; height: 20px;">10</td><td style="width: 20px; height: 20px;">11</td><td style="width: 20px; height: 20px;">12</td><td style="width: 20px; height: 20px;">13</td><td style="width: 20px; height: 20px;">14</td><td style="width: 20px; height: 20px;">15</td><td style="width: 20px; height: 20px;">16</td><td style="width: 20px; height: 20px;">17</td><td style="width: 20px; height: 20px;">18</td><td style="width: 20px; height: 20px;">19</td><td style="width: 20px; height: 20px;">20</td> </tr> </table> </div> <div style="display: flex; justify-content: center; align-items: center; gap: 20px; margin-bottom: 10px;"> <div style="border: 1px solid gray; padding: 5px; background-color: #ffff00;">5</div> <div style="border: 1px solid gray; padding: 5px; background-color: #90ee90;">7</div> </div> <div style="font-size: 48px; font-weight: bold; text-align: center;">5 + 7 = 12</div> <div style="text-align: center; margin-top: 10px;">  <p style="font-size: 12px;">Correct!</p> </div>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	<p>You can solve this problem anyway you want. Using your head, using your fingers or using the ten frame. When you have your answer press the "Ready to Answer" button.</p> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> </div> <div style="text-align: center;">  <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> </div> </div> <div style="font-size: 48px; font-weight: bold; text-align: center; margin-top: 20px;">8 + 3</div> <div style="text-align: center; margin-top: 20px;"> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0; display: inline-block;">Ready To Answer</div> <div style="border: 1px solid gray; padding: 5px; background-color: #f0f0f0; display: inline-block; margin-left: 20px;">Start Over</div> </div>																
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																		
<div style="display: flex; justify-content: space-around; align-items: center; margin-bottom: 10px;"> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> </div> <div style="font-size: 48px; font-weight: bold; text-align: center;">8 + 3 = 11</div> <div style="text-align: center; margin-top: 10px;">  <p style="font-size: 12px;">Correct!</p> </div>																	<div style="text-align: center; margin-bottom: 20px;"> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> <table border="1" style="margin: 0 auto; border-collapse: collapse;"> <tr><td></td><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td></tr> </table> </div> <div style="font-size: 48px; font-weight: bold; text-align: center;">8 + 3 = 11</div>																					

Appendix H continued

<p>How much is four and eight more?</p> <p>You can figure this out anyway you want, you can use your head, you can use your fingers you can count using the dots in the ten frames. When you're ready enter your answer using the number pad below and hit done.</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>4</p> </div> <div style="font-size: 2em;">+</div> <div style="text-align: center;">  <p>8</p> </div> </div> <div style="display: flex; justify-content: center; margin-top: 10px;"> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">1</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">3</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">5</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">7</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">9</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">Done</div> </div> <div style="display: flex; justify-content: center; margin-top: 10px;"> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">0</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">2</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">4</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">6</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">8</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">Delete</div> </div>	<div style="text-align: center;">  <p>4 + 8 = 12</p> </div> 
<p>You can use your fingers, you can use head, or click on the number of X to cross out 8 of the 12 dots. Enter your answer on the numberpad below.</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  </div> <div style="text-align: center;"> <p>12 - 8</p>   </div> </div> <div style="display: flex; justify-content: center; margin-top: 10px;"> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">1</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">3</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">5</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">7</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">9</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">Done</div> </div> <div style="display: flex; justify-content: center; margin-top: 10px;"> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">0</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">2</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">4</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">6</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">8</div> <div style="border: 1px solid gray; padding: 2px 5px; margin: 2px;">Delete</div> </div>	<div style="text-align: center;"> <p>4 + 8 = 12</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> part part whole </div> </div> <div style="text-align: center; margin-top: 20px;">  </div> <div style="text-align: center; margin-top: 20px;"> <p>12 - 8 = 4</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> whole part part </div> </div>
<div style="text-align: center;"> <p>4 + 8 = 12</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> part part whole </div> </div> <div style="text-align: center; margin-top: 20px;">  </div> <div style="text-align: center; margin-top: 20px;"> <p>12 - 8 = 4</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> whole part part </div> </div>	<div style="text-align: center;"> <p>4 + 8 = 12</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> part part whole </div> </div> <div style="text-align: center; margin-top: 20px;">  </div> <div style="text-align: center; margin-top: 20px;"> <p>12 - 8 = 4</p> <div style="display: flex; justify-content: space-around; font-size: 0.8em;"> whole part part </div> </div>

Appendix H continued

<p>$6 + 6 = ?$</p> 	<p>six and six more is twelve</p> <p>$6 + 6 = 12$</p> <p>part part whole</p> 
<p>Does six and six more is twelve help you to answer...</p> <p>$6 + 6 = 12$</p> <p>part part whole</p>  <p>$12 - 6 = ?$</p>	<p>Does 6 and 6 more is 12 help you to answer what 12 take away 6 is?</p> <p>No, because they do NOT have the same whole and the same part.</p> <p>Yes, because they have the same whole and the same part.</p>  <p>$12 - 6 = ?$</p> <p>whole part part</p>
<p>twelve take away six is six</p>  <p>$12 - 6 = 6$</p> <p>whole part part</p>	<p>Wonderful, five stars! Yes is correct because when the part six and the part six makes the whole twelve, the whole twelve take away the part six leaves the part six.</p> <p>$6 + 6 = 12$</p> <p>part part whole</p>  <p>$12 - 6 = 6$</p> <p>whole part part</p>

Appendix H continued

<p>Ready?</p> <p>If you do not know the answer right away, make a smart guess as fast as you can.</p>	<p>How much is nine and nine more?</p> $9 + 9$
<p>How much is nine and nine more?</p> $9 + 9$ <div><div>13579</div><div>02468</div><div>Done</div><div>Delete</div></div> <p>18</p>	<p>Awesome!</p> <p>★★★★★</p> $9 + 9 = 18$

Appendix I: False Positive Identification

1. Response bias determination were done separately for each session and based on all items given in a session. Addition and subtraction items were tabulated separately. This was done because it is not uncommon for children to change a predominant strategy in a later session (e.g., adopt a more advanced strategy) or to tailor a response to a particular type of problem (e.g., stating the second number [minuend] on subtraction problems only). The determination was done for the two sets presented in a session instead of by one set to insure a sufficient number of examples of a response bias to make a reliable determination.
2. Scorers looked for common addition response biases such as *stating a favorite number* (e.g., typically responding to various items with “five”) or *make-a-teen* (e.g., responding with *n*-teen using one of the addends) and common subtraction response biases such as *state-the-minuend-or-the subtrahend* (e.g., responding with either the minuend or the subtrahend), *state-the-subtrahend* (e.g., consistently responding with the subtrahend) or *stating a favorite number*.
3. If a single response bias was apparent, then a scorer scored the response to each item in one of five ways:
 - a. Possible inappropriate use of a strategy (resulting in an incorrect answer);
 - b. Possible appropriate use of a strategy (resulting in a correct answer);
 - c. Other incorrect estimates (error not attributable to the strategy);
 - d. Other correct estimates (correct response not attributable to the strategy); and
 - e. Use of concrete or abstract counting strategy or no response.
4. If two (or more) response biases were apparent, then the scorer scored the response to each item for each separately using the categories listed in guideline #3. The more prevalent strategy was used for further analyses.
5. In order to rise to the level of a response bias, four criteria had to be met:
 - a. Total inappropriate uses is half or more of all estimation errors: $a_{TOT} \div (a_{TOT} + c_{TOT}) \geq .50$;
 - . Total number of inappropriate uses is more 25% of all trials¹: $a_{TOT} \geq .25$ (minimum requirement 6 addition items or 4 subtraction items) and $a_{TOT} \div N_{adj} \geq .25$ (where $N_{adj} = N - \text{number of items meeting category 3e}$);
 - a. Total uses of strategy is half or more of all estimates: $(a_{TOT} + b_{TOT}) \div (a_{TOT} + b_{TOT} + c_{TOT} + d_{TOT}) \geq .50$; and
 - b. Total uses of strategy is half or more of all trials: $(a_{TOT} + b_{TOT}) \div N \geq .50$ (minimum of $N \div 2$ items).

These criteria further ensured a consistency of strategy use over a variety of items and a sufficient number of trials to ensure a reliable determination of a response bias. Counted or no responses are excluded from the tabulation for points 5a, 5b, and 5c.

Note¹. The criterion for inappropriate uses is more than 50% on subtraction given the limited number of items per session.

Appendix I continued

Subtraction Example 1:

Item	Answer	Strategy	Score (Guideline # 3)	a _{TOT}	b _{TOT}	c _{TOT}	d _{TOT}
12 - 9	9,6	Spontaneous					
10 - 5	5	Correction	c. Other incorrect estimates (error not attributable to the strategy)			1	
11 - 8	8	Other	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
14 - 7	7	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
13 - 4	4	Other	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
12 - 5	4	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
12 - 7	7	Other	c. Other incorrect estimates (error not attributable to the strategy)			1	
		Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			

Subtraction Example 2:

Item	Answer	Strategy	Score (Guideline # 3)	a _{TOT}	b _{TOT}	c _{TOT}	d _{TOT}
12 - 8	8	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
11 - 7	6	Automatic	c. Other incorrect estimates (error not attributable to the strategy)			1	
10 - 6	6	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
10 - 7	7	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
12 - 6	6	Automatic	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
11 - 4	4	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
13 - 9	9	Automatic	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			

Response Bias Criterion	Example 1	Example 2
a. $a_{TOT} \div (a_{TOT} + c_{TOT}) \geq .50$	$3 \div (3 + 2) = .60$	$5 \div (5 + 1) = .83$
b. $a_{TOT} \geq 4$ and $a_{TOT} \div N_{adj} \geq .25$	3 trials ; $3 \div 7 = .42 \geq .25$	5 trials ; $5 \div 7 = .71 \geq .25$
c. $(a_{TOT} + b_{TOT}) \div (a_{TOT} + b_{TOT} + c_{TOT} + d_{TOT}) \geq .50$	$(3 + 2) \div (3 + 2 + 2 + 0) = .71$	$(5 + 1) \div (5 + 1 + 1 + 0) = .86$
d. $(a_{TOT} + b_{TOT}) \div 7 \geq .50$ (minimum requirement 4 trials)	$(3 + 2) \div 7 = .71$	$(5 + 1) \div 7 = .86$
Response bias?	None – Criterion b not met.	State-the-minuend strategy

Appendix I continued

Addition Example

Item	Answer	Strategy	Score (Guideline # 3)	a _{TOT}	b _{TOT}	c _{TOT}	d _{TOT}
9 + 5	15	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
7 + 9	17	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
10 + 4	14	Automatic	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
7 + 6	16	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
4 + 7	10	Other	c. Other incorrect estimates (error not attributable to the strategy)			1	
6 + 10	16	Automatic	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
8 + 5	18	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
7 + 8	11	Other	c. Other incorrect estimates (error not attributable to the strategy)			1	
5 + 6	NR	No Response	e. Use of concrete or abstract counting strategy or no response	-	-	-	-
8 + 6	16	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
3 + 10	13	Other	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
8 + 7	16	Other	c. Other incorrect estimates (error not attributable to the strategy)			1	
10 + 7	17	Automatic	b. Possible appropriate use of a strategy (resulting in a correct answer)		1		
5 + 4	13	Other	c. Other incorrect estimates (error not attributable to the strategy)			1	
9 + 7	17	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			
7 + 5	NR	No Response	e. Use of concrete or abstract counting strategy or no response	-	-	-	-
6 + 7	17	Other	a. Possible inappropriate use of a strategy (resulting in an incorrect answer)	1			

Guideline #5: Four criteria for a *make-a-teen strategy* were met:

- $a_{TOT} \div (a_{TOT} + c_{TOT}) \geq .50 = \text{YES: } 7 \div (7 + 4) = .64;$
- $a_{TOT} \geq 6 \text{ and } a_{TOT} \div N_{adj} \geq .25 = \text{YES: } 7 \geq 6; 7 \div 17 = .41 \geq .25$
- $(a_{TOT} + b_{TOT}) \div (a_{TOT} + b_{TOT} + c_{TOT} + d_{TOT}) \geq .50 = \text{YES: } (7 + 4) \div (7 + 4 + 4 + 0) = .73; \text{ and}$
- $(a_{TOT} + b_{TOT}) \div N \geq .50 \text{ (minimum requirement 10 trials)} = \text{YES: } (7 + 4) \div 17 = .65.$