STRATEGY USE AND BASIC ARITHMETIC COGNITION IN ADULTS

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Psychology

University of Saskatchewan

Saskatoon

By

Arron W. S. Metcalfe

PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Psychology
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A5
Canada

ABSTRACT

Arithmetic cognition research was at one time concerned mostly with the representation and retrieval of arithmetic facts in memory. More recently it was found that memory retrieval does not account for all single digit arithmetic performance. For example, Canadian university students solve up to 40% of basic addition problems using procedural strategies (e.g. 5 + 3 = 5 + 1 + 1 + 1). Given that procedures are less efficient than direct memory retrieval it is important to understand why procedure use is high, even for relatively skilled adults. My dissertation, therefore, sought to expand understanding of strategy choice for adults' basic arithmetic. Background on this topic and supporting knowledge germane to the topic are provided in Chapter 1.

Chapter 2 focused on a well-known, but unexplained, finding: A written word problem (six + seven) results in much greater reported use of procedures (e.g., counting) than the same problem in digits (6 + 7). I hypothesized that this could be the result of a metacognitive effect whereby the low surface familiarity for word problems discourages retrieval. This was tested by familiarizing participants with a subset of the written word stimuli (e.g. three + four = ?, six + nine= ?) and then testing them on unpractised problems comprised of practiced components (four + six = ?). The result was increased retrieval reported for unpractised problems with practiced components. This indicates that surface familiarity contributes to strategy choice.

Chapter 3 focused on another classic phenomenon in the arithmetic cognition literature, the problem size effect: Response time, error, and procedure rates increase as a function of problem size. A previous study reported a reduced problem size effect for auditory multiplication problems compared to digit problems. I hypothesized that if this reduction was due to problem encoding processes rather than an effect on calculation per se, then a similar pattern would be observed for addition. Instead, I found that the size effect for addition was larger. I concluded that the auditory format promotes procedures for addition, but promotes retrieval for multiplication.

Chapters 4 and 5 were concerned with a well-known methodological issue in the strategy literature, subjectivity of self-reports: Some claim self-reports are more like opinions than objective measures. Thevenot, Fanget, and Fayol (2007) ostensibly solved this problem by

probing problem memory subsequent to participants providing an answer. They reasoned that after a more complex procedure, the memory for the original problem would become degraded. The result would be better memory for problems answered by retrieval instead of by procedure. I hypothesized that their interpretation of their findings was conflated with the effect of switching tasks from arithmetic to number memory. I demonstrated that their new method for measuring strategy choice was contaminated by task switching costs, which compromises its application as a measure of strategy choice (Chapter 4). In a subsequent project (Chapter 5), I tested the sensitivity of this new method to detect the effects of factors known in the literature to affect strategy choice. The results indicated that Thevenot et al.'s new method was insensitive to at least one of these factors. Thus, attempts to control for the confounding effects of task switching described in Chapter 4, in order to implement this new measure, are not warranted.

The current dissertation expanded understanding of strategy choice in four directions by 1) demonstrating that metacognitive factors cause increases in procedure strategies, 2) by demonstrating that the process of strategy selection is affected differentially by digit and auditory-verbal input, 3) by investigating the validity of an alternative measure of strategy use in experimental paradigms, and 4) by discovering a critical failure in the sensitivity of this new measure to measure the effects of factors known to influence strategy use. General conclusions are discussed in Chapter 6.

DEDICATION

I would like to dedicate this dissertation to my wife Carrie, our acquaintance began about the same time as my graduate program of study and she has been there every step of the way. I don’t know if I would have finished without her constant and unwavering support. In turn, I also thank this document for my love and marriage to this woman, for without the circumstances related to the inception of my dissertation I would never have met her.

ACKNOWLEDGMENTS

I could not have completed this dissertation without the supportive presence of my adviser Jamie Campbell, who taught me everything I didn't know about academia, writing, and arithmetic cognition. I would be fortunate to become even half the researcher this man is.

The time and effort of my dissertation committee was much appreciated. Valerie Thompson, Ron Borowsky, and Shaun Murphy each contributed in their own unique way. Without their contribution many useful insights would have been missed.

The research in this dissertation and my tenure as student may not have been possible without the continuing support of the Natural Sciences and Engineering Research Council of Canada and their various student funding initiatives that I count to be among the best on Earth.

I owe a large part of the knowledge and techniques underlying my expertise in cognitive psychology to the many professors that contributed to my course work over the years. Though I could not possibly name all of these mentors, Jamie Campbell, R. Borowsky, and V. Thompson are most assuredly included. As well, I would add Professors Oleson, Edguer, and Corenblum whom were among my most difficult undergrad teachers and always challenged me to do more. I would also like to thank Dr. Barbara Gfellner who originally inspired me to be a researcher by giving me those all-important first lab opportunities in third year undergrad.

I would like to recognize the support I have gotten from the Cognitive Science Lab at the University of Saskatchewan over the years, especially fellow grad students including Jamie Prowse Turner, David Lane, Erin Beatty, and Greg Krätzig. The value of their help and advice cannot be overstated. The commiserating and late Friday afternoons were instrumental to my function as well. This list also includes many other grad students throughout the department too numerous to name. The Lab Administrators always had a helping hand, including J. Shynkaruk, N. Robert, M. Bayly, L. Aspenleider, D. Canales, G. Pennycook, and S. Sacher; I am also appreciative of the contributions of numerous RAs that have worked with the lab over the years.

Finally, I would like to thank my family, also too numerous to name, but without which I would have not had the formative experiences that make part of who I am. Although as we all grow older I find we grow further apart, I always keep my early reminiscences close at hand.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASCM	Adaptive Strategy Choice Model
HF	high familiarity
LF	low familiarity
ms	milliseconds
ORP	Operand Recognition Paradigm
PSE	problem size effect
RT	response time
s	seconds

CHAPTER 1

Strategy Use and Basic Arithmetic Cognition in Adults

Arithmetic includes the mathematical operations of addition, subtraction, multiplication, and division, it is the oldest branch of mathematics, and is the most commonly used set of mathematical operations (Carnahan, 1946). Furthermore, numeracy – the ability to reason with numbers – has been linked to occupational attainment, income, and quality of life indicators such as body mass index, coping ability in chronic illness, and prognosis in terminal illness (Apter et al., 2006; Charette & Meng, 1998; Finnie & Meng, 2006; Huizinga, Beech, Cavanaugh, Elasy, & Rothman, 2008; Waldrop-Valverde et al., 2009). Thus, understanding the mental processes of adult arithmetic is of vital importance for reasons ranging from an intrinsic interest in the most ubiquitous and ancient of maths, to the application of that understanding for practical use.

Despite its importance, arithmetic was historically one of the more ignored domains of cognitive psychology. Indeed, as an example of the paucity of interest in the area circa 1970, Ashcraft (1992) commented that the sum of our knowledge on the topic was that counting backward by threes was a superlative distracter task for short-term memory[1]. Since then the field of inquiry has developed and expanded. By the early 1990s a dissertation on arithmetic cognition would have supplied the reader with sophisticated associative memory models that explained arithmetic performance as a function of the strength of connections amongst problems and facts as built by experience (e.g., Ashcraft, 1992; Campbell, 1995; McCloskey, Harley, & Sokol, 1991). These models had a heavy emphasis on retrieval based processes predicated on semantic long-term memory. Although successful in terms of fitting data, and useful in terms of explanatory power, these models have been found to be incomplete due to their exclusive emphasis on long-term memory.

More recently, it has been found that memory retrieval does not account for all, or even a necessarily large majority of adults' answers to basic arithmetic problems (e.g., LeFevre et al., 1996b; LeFevre, Sadesky, & Bisanz, 1996). For example, it has been demonstrated that Canadian

---

[1] So useful that it is still commonly utilized today, see for example Campbell, Dowd, Frick, McCallum, & Metcalfe (2010).

university students solve up to 40% of some basic addition problems using procedural strategies[2] (Campbell & Xue, 2001). Given that procedures are often less efficient than direct memory retrieval in terms of speed and accuracy, it is important to understand why procedure use persists, even in relatively skilled adults that have arguably rehearsed the basic arithmetic facts thousands of times in the course of their educational and personal histories.[3]

Thus, the thesis of this dissertation is to understand strategy use for basic arithmetic. The dissertation is organized in a manuscript by chapter format and the experiments that follow in Chapters 2-5 are best understood in the context that they are independent works published in academic journals (Chapters 2-4), or in the case of Chapter 5, an article under review. Consequently, although each project in this dissertation was designed to illuminate an aspect of strategy use in basic adults' arithmetic cognition, each project does not necessarily follow logically from the next. Instead, at the beginning of each chapter, an editorial bridge will explain the significance of the work in relation to my primary thesis. Citations for the original publication of these manuscripts also appear at the beginning of each chapter. Within each chapter, individual experiments are numbered consecutively from Experiment 1. Chapter 6 will finish this dissertation by explaining the relevance of the findings from Chapters 2-5 to the field of arithmetic cognition as a whole. Detailed theoretical and empirical background relevant to each chapter is developed within the chapter in question. An overview, however, is also provided here.

Chapter 2 investigated the causal role of cue familiarity and its effect on the prevalence rates of direct retrieval versus procedural strategies. Specifically, the focus was on a well-known but unexplained finding: A written word problem (e.g., six + seven) results in much greater reported use of procedures (e.g., counting) than the same problem in Arabic digit format (e.g., 6 + 7). I hypothesized that this could be the result of a metacognitive effect whereby the low surface familiarity for word problems discourages retrieval. This was tested by familiarizing participants with a subset of the written word stimuli (e.g., three + four = ? and six + nine = ?)

---

[2] So-called for the set of steps used to solve the problem, e.g., counting: $5 + 3 = 5 + 1 + 1 + 1 = 8$; transformation: $17 – 9 = 17 – 10 = 7 + 1 = 8$.

[3] At least up to 9×9 in multiplication and arguably up to $9 + 9$ for addition as well.

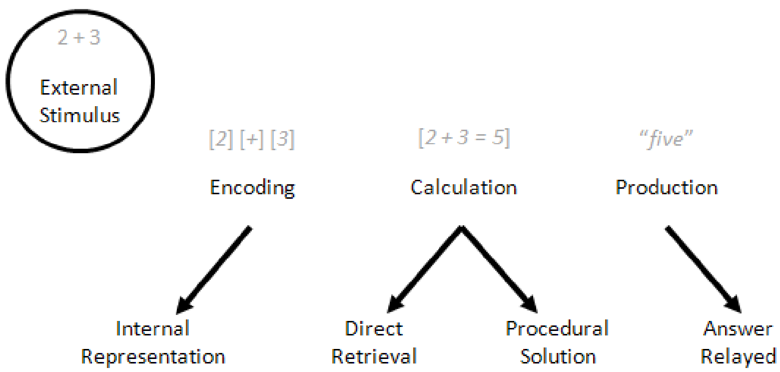
and then testing them on unpractised problems comprised of the familiarized operands (e.g., four + six = ? and three + nine). The result was increased retrieval reported for unpractised problems with familiarized operands versus unpractised problems with unfamiliarized operands (i.e., respectively, four + six vs. two + eight). This indicates that surface familiarity contributed to strategy choice.

Similarly, Chapter 3 investigated another classic phenomenon, the *problem size effect*. This refers to the finding that response time, errors, and procedure rates all increase as a function of problem size (usually defined by the magnitude of the quantities in the problem or answer). A previous study (LeFevre, Lei, Smith-Chant, & Mullins, 2001) reported a reduced problem size effect for multiplication problems presented in auditory-verbal format (i.e., spoken words) compared to problems presented in Arabic-digit format (e.g., 2×4). I hypothesized that if this reduction was due to problem encoding processes rather than an effect on calculation per se, then a similar pattern would be observed for addition. Instead, I found an inverted size effect for addition, that is, a larger size effect for auditory-verbal format than Arabic-digit format (i.e., a three-way interaction). Decomposition of the RT distribution in to the main area of the distribution (μ) and the exponentially distributed upper tail (τ), demonstrated that auditory format presentation resulted in a faster mean on the μ distribution than Arabic format, but that Arabic format resulted in a faster mean on the τ distribution than auditory presentation. These results demonstrated that procedural strategies (primarily associated with τ) are facilitated by Arabic input and that retrieval strategies (primarily associated with μ) are promoted by auditory input. This unique behavioural evidence dovetailed nicely with neuro-cognitive evidence of a dissociation between storage and retrieval of verbally rehearsed facts (e.g. “nine times nine equals four”) associated mostly with multiplication and the solution of problems via procedures predicated upon number specific analogical representations associated mostly with subtraction and a large minority of addition problems (e.g. adding number specific magnitude quantities on an internalized number line representation).

A corollary to understanding strategy use for single-digit by single-digit arithmetic problems is the measurement of strategy use in basic arithmetic task performance, thus, a portion of this dissertation is devoted to this question. One common measure of strategy use has been the verbal self-report. In this method, participants are asked to explain how a problem was answered or to choose the strategy that was used from a list of possible strategies (e.g., remember, transform, count, other/don't know). Although this method has appeared in dozens of studies (e.g., Campbell & Albert, 2009; Campbell & Austin, 2002; Campbell & Fugelsang, 2003; Campbell & Gunter, 2002; Campbell & Penner-Wilger, 2006; Campbell & Timm, 2000; Campbell Xue, 2001; Geary & Wiley, 1991; Hecht, 1999; Imbo, Vandierendonck, & Rosseel, 2007; LeFevre et al., 1996b; LeFevre, Sadesky, & Bisanz, 1996) it has been criticized as subjective and inaccurate (Kirk & Ashcraft, 2001; Thevenot, Fanget, & Fayol, 2007). These criticisms have recently led to the proposal of a new method to measure strategy use whereby the memory strength of a problem immediately after solving the problem is proposed to be an index of strategy. This method is known as the *operand recognition paradigm*. Whereas direct retrieval is simple (e.g., 7 + 2 = 9), it preserves memory strength for the problem operands (i.e., for 7 and 2); whereas procedures are complex (e.g., 28 + 13 = 28 + 10 = 38 + 3 = 41), this complexity reduces relative memory strength of problem operands (i.e., for 28 and 13) via interference from the intermediate steps (Thevenot, Barrouillet, & Fayol, 2001; Thevenot, Castel, Fanget, & Fayol, in press; Thevenot et al., 2007). Chapter 4 establishes that this new method touted to objectively measure strategy use is confounded by the cognitive cost of switching between arithmetic problem solving and the competing cognitive task set of number matching (i.e., the proposed method of strategy measurement) in memory. However, due to the possible importance of this method to the field of arithmetic cognition, further work was undertaken to establish if the method was worth modifying or correcting. Thus Chapter 5 quantifies the relationship between factors known to affect strategy use for simple arithmetic and the effect of these factors on the new operand recognition measure, and directly compares these to results from verbal self-reports.

The following research on strategy use in basic arithmetic for adults will be best understood within the superordinate context of the general process of arithmetic cognition. Thus I provide to the reader here a standard model of arithmetic cognition for this purpose. In the generic and widely assumed three stage model, the process is composed of three parts: The encoding stage, the calculation stage, and the production stage (Campbell & Epp, 2005). As depicted in Figure 1-1, an arithmetic problem begins as an external stimulus that is converted into an internal representation including the operands and the operation to be performed.

Figure 1-1. Generic Model of Arithmetic Processing



Note. Black = stages of processing. Grey = arithmetic representations. [*x*] = generic mental representation. Function of stages indicated with arrows on bottom tier. Processing proceeds from left to right. Based on description from Campbell and Epp (2005).

During the calculation stage, the answer to the problem is either remembered or computed by some form of elaborated decomposition or serial addition or subtraction. Finally, the answer to the problem is converted into the appropriate production output which can include verbal, written, or verification formats. The work in the upcoming chapters concerns itself mainly with the encoding and calculation stages within the standard model. As discussed in Campbell and Epp (2005), specific models diverge on assumptions ranging from the format of the encoded representation (e.g., abstract, verbal, or analogue) to whether or not each stage is affected by the input format of the previous stage. These assumptions and the data that supports these assumptions are, however, beyond the scope of this introduction but will be explicated when specifically relevant in each of the following chapters. For a good and thorough review of this work I direct the reader to Campbell and Epp (2005).

In summary, the current dissertation expands understanding of strategy choice in three primary domains by 1) demonstrating that metacognitive factors affect strategy choice for basic arithmetic, 2) by demonstrating that the process of strategy selection is affected differentially by Arabic-digit and auditory-verbal input which also interacts with the operation to be solved, and 3) by discovering critical limitations in a new paradigm that has begun to gather momentum as an important measure of strategy use for arithmetic. Collectively this work represents a robust and diverse contribution to the literature on arithmetic strategy use. Finally, readers should keep in mind that the topical diversity of this research program was specifically undertaken to build a strong grounding for future work in multiple research streams around the core topic of arithmetic cognition and should be interpreted thusly.

CHAPTER 2

Are adults' decisions to use direct memory retrieval for simple addition influenced by the familiarity of problem operands? Studies consistently demonstrate that educated adults do not exclusively use direct retrieval to answer even the most basic arithmetic problems such as 2 + 5. This is especially true when these problems are presented in formats other than standard Arabic digit format, such as written word format (e.g., two + five = ?). Although many causes for this phenomenon have been explored in past research, a complete explanation has not yet been found. Here, we manipulated the familiarity of a subset of operands by having adults repeatedly practice specific additions (two + five = ?; Experiment 1) or magnitude comparisons (two five, choose the larger; Experiment 2). Both experiments provided evidence that pre-exposure to single-digit operands increased reported use of direct retrieval for new combinations of the familiarized operands. RT and error patterns across experiments also supported the conclusion that increased use of retrieval facilitated performance. These results show that operand familiarity potentially plays a significant role in adults' strategy choices for simple addition. Thus, current and future models of basic arithmetic need to account for this previously unconsidered metacognitive factor when describing adults’ basic arithmetic solution processes.

This chapter has been previously published in a Taylor & Francis journal:

Metcalfe, A. W. S., & Campbell, J. I. D. (2007). The role of cue familiarity in adult’s strategy choices for simple addition. *European Journal of Cognitive Psychology, 19*, 356-373.

Cue Familiarity and Strategy Choice

Calculation skills have provided an important domain for studying the theoretical relation between fact-retrieval and procedural skills (Logan & Klapp, 1991; Rickard, 2005; Sohn &

Carlson, 1998; Zbrodoff, 1995). Use of fact-retrieval vs. procedural strategies for calculation is also of practical importance. Comparisons of simple arithmetic performance of Asian and North American university students indicate that cross-national differences favouring Asian populations are due, in part, to North-American adults' relatively poor memory for simple arithmetic (e.g., 6 + 8 = 14, 7 × 9 = 63) and relatively greater reliance on less-efficient procedural strategies such as counting (Campbell & Xue, 2001; Geary, 1996; LeFevre & Liu, 1997). Understanding the cognitive mechanisms that mediate arithmetic strategy choice contributes to understanding variability in mathematical competence as well as to understanding the relationship between fact-retrieval and procedural memory processes. The following experiments investigated the role of operand familiarity in adults' use of fact-retrieval versus procedural strategies for simple addition.

Direct memory retrieval appears to be the dominant strategy for simple addition reported by educated adults in Canada or the U.S., but numerous experimental studies indicate substantial use of procedures such as counting or transformation (e.g., 6 + 7 = 6 + 6 + 1 = 13). Procedure use is reported on about 20% to 30% of simple addition trials (21% in Campbell & Austin, 2002; 24% for non-Asian Canadians in Campbell & Xue, 2001; 31% in Campbell & Timm, 2000; 27% in Geary, 1996; 34% in Hecht, 1999; 29% in LeFevre, Sadesky, & Bisanz, 1996). Several variables are known to influence adults' strategy choices for simple arithmetic (Campbell & Gunter, 2002). For example, direct memory retrieval is much more likely for numerically small problems (e.g., 2 + 3) compared to large problems (9 + 7) (Campbell & Xue, 2001; Kirk & Ashcraft, 2001; LeFevre et al., 1996a; Seyler, Kirk, & Ashcraft, 2003). This problem-size effect on retrieval usage is typically explained in terms of greater memory strength for small problems than for large problems (Siegler & Shipley, 1995). More generally, theoretical models of strategy choice for simple arithmetic emphasize the role of the strength of the association between problem elements (i.e., the two operands and the answer) in determining retrieval probability. The greater the memory strength, the greater the probability that a problem is solved by direct memory retrieval rather than a procedural strategy (LeFevre et al., 1996a; Rickard, 2005; Siegler & Lemaire, 1997; Siegler & Shipley, 1995).

Another factor that affects strategy choice is the surface format of the problem operands (e.g., Arabic digits vs. written number words). Campbell, Parker and Doetzel (2004) found that

Canadian university students were 50% more likely to report procedural strategies (e.g., counting, transformation) for simple addition with written English operands (four + eight) compared to Arabic digits (4 + 8) (cf. Campbell & Fugelsang, 2001). Why would operands in word format disrupt retrieval and promote use of procedures? According to Siegler and Shipley's (1995) *Adaptive Strategy Choice Model* (ASCM) selection of an arithmetic strategy depends on its relative efficiency (i.e., speed and probability of success). Consequently, manipulations that reduce retrieval efficiency promote a switch to procedures (Campbell & Timm, 2000). For example, one way that the word format interferes with performance is by promoting operand-intrusion errors (e.g., 8 × 4 = 24, 2 + 9 = 9) where one of the operands appears in the error. Operand intrusions are much more likely with word than digit stimuli, particularly for numerically larger problems (Campbell, 1994; LeFevre, Lei, Smith-Chant, & Mullins, 2001; Noël, Fias, & Brysbaert, 1997). This word-specific interference potentially reduces the efficiency of retrieval relative to digit problems, which could promote a switch from retrieval to procedures according to the ASCM model.

Nevertheless, there are other factors that influence strategy selection for arithmetic that raise the possibility that retrieval strength or efficiency may not be the whole story. Specifically, familiarity with problem operands could be an important factor in choosing to retrieve rather than to calculate (Reder & Ritter, 1992; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997). Schunn et al. showed that attempted use of direct retrieval for newly learned complex arithmetic facts (52 × 34 = ?) increased with the familiarity of problem operands independently of the availability of the answer. This finding can be explained by the *source of activation confusion model* of memory retrieval (cf. Schunn et al., 1997). In this view, repeatedly practicing a problem strengthens associations with the problem operands. These associations are activated when a new problem involving those operands is encountered. The activation of previously strengthened associations produces a feeling of knowing the answer that results in a decision to attempt retrieval rather than to calculate a solution.

Similarly, Koriat, and Levy-Sadot (2001) have proposed that *cue-familiarity* (i.e., the familiarity of components of a problem or question), influences the decision to attempt direct memory retrieval. In their model, increased cue-familiarity of a problem promotes retrieval attempts, although accessibility or memory strength may also play a role. Koriat and Levy-

Sadot's model potentially explains why Schunn et al. (1997) observed decisions to retrieve based on cue-familiarity despite participants' lack of knowledge of the problem's answer. Furthermore, the influence of cue-familiarity on decisions to retrieve is stronger under conditions of time pressure (Benjamin, 2005). As time pressure is a common feature of experiments on simple arithmetic, cue familiarity could play an important role in people's strategy choices for simple arithmetic under experimental conditions.

Given the evidence that cue familiarity can affect strategy choice in other domains, it is worthwhile to determine if operand familiarity can similarly influence retrieval usage for elementary arithmetic facts. Differences in familiarity could explain the finding of Campbell et al. (2004) that direct retrieval was much more likely when simple addition problems appeared in digit format (4 + 8) compared to written-word format (four + eight). Arithmetic problems are frequently encountered in Arabic format, but rarely encountered in written word format. Consequently, the relatively low familiarity of written-word arithmetic problems could discourage use of direct memory retrieval. Currently, however, none of the theoretical models concerning strategy choice for simple arithmetic have considered the role of operand familiarity in strategy choice, independent from the strength of the problem – answer association.

In the following experiment, we tested Canadian university students on simple addition problems (two + five, seven + eight, etc.), and asked them to report their strategy after each trial by selecting from Remember, Count, Transform or Other. Theoretically, selection of the Remember category corresponds to direct memory retrieval, whereas selection of one of the other categories corresponds to procedure use (Campbell & Austin, 2002; Campbell & Fugelsang, 2001; Campbell et al., 2004; Campbell & Xue, 2001). Participants received a practice phase in which they repeatedly solved a subset of addition problems, which familiarized four of the eight addends between 2 and 9. They were then tested on unpractised addition problems comprised of familiarized or unfamiliarized operands. If greater operand familiarity promotes the use of direct memory retrieval, then we would expect greater reported use of retrieval for the high familiarity problems compared to the low familiarity problems.

Experiment 1

Method

Participants. Sixty introductory Psychology students (56 female, 4 male) participated for course credit at the University of Saskatchewan. The mean age of participants was 19.0 years, 55 participants were right handed, and 45 participants claimed English as their first language for arithmetic.

Stimuli and Design. Stimuli were single-digit addition problems displayed in lower case English word format. We used word format rather than digit format because retrieval usage is lower with words than digits (Campbell et al., 2004), which should provide the best opportunity to observe increased use of retrieval as a function of increased operand familiarity. Problems appeared on a computer monitor as white characters against a black background. Characters were approximately 4 mm wide and 7 mm high; the addition symbol bounded by one space on each side separated each word.

The experiment was composed of two phases: a practice phase and a test phase. The order of problems during both the practice and test phases was independently randomized for each participant. During the practice phase, participants repeatedly practiced a set of simple addition problems in order to become especially familiar with a subset of four operands. The practice set consisted of eight problems composed either from the operands 2, 5, 7, 8 or 3, 4, 6, 9. Even-numbered participants practiced two + five, five + two, seven + eight, eight + seven, two + two, five + five, seven + seven, eight + eight, and odd-number participants practiced three + four, four + three, six + nine, nine + six, three + three, four + four, six + six, and nine + nine. Six randomized orders of the eight practice problems appeared in a single block of trials. During the test phase, participants received four blocks of 36 simple addition problems. The problems involved all possible single-digit pairs of addends from 2 to 9. Approximately half of the non-tie problems (e.g., two + three and three + two) were randomly assigned to be tested in one order or the other in the first block; operand order then alternated across blocks.

Procedure. The following instructions were displayed on the computer screen at the beginning of the experiment:

> There will be 4 blocks of 36 simple additions involving single-digit plus single-digit problems. Problem format will be English words (one + two). After each problem, please indicate how you solved the problem by choosing from among the following strategies... Transform, Count, Remember, Other. Select TRANSFORM if you used knowledge of a related problem. Select COUNT if you used a strategy based on counting. Select REMEMBER if the answer seemed to come to you without any intermediate steps, inferences, or calculations. Select OTHER if you used some other strategy or are uncertain. A fixation dot will flash twice before each problem appears. Please respond as quickly and accurately as possible. Occasional errors are normal and should not concern you. To warm up, there will be a block of 48 practice trials. If you proceed quickly, testing will take only about 25 minutes.

Verbal instructions advised the participant to concentrate on each problem in the practice phase, and informed them that their responses would be timed. The experimenter also read through a paper copy of the strategy definitions, which the participant retained throughout the experiment. The paper copy of the definitions read as follows:

> TRANSFORM: You solve the problem by referring to a related problem in the same or another operation. For example, you might solve 17 – 9 =? by remembering that 17 – 10 = 7, so 17 – 9 must equal 8. COUNT: You solve the problem by counting a certain number of times to get the answer. REMEMBER: You solve the problem by just remembering or knowing the answer directly from memory without any intervening steps. OTHER: You may solve the problem by a strategy unlisted here, or you may be uncertain how you solved the problem.

The experimenter initiated each block when the participant indicated they were ready. For each practice and experimental trial a fixation dot flashed twice over a 1 s interval before being replaced by the problem with the plus sign (+) at the center of the screen. Timing began when the problem appeared on the screen. At the onset of the participant's response, a voice-activated switch terminated the clock via a microphone worn by the participant. Timing was accurate to within ± 1 ms. The response caused the problem to disappear from the screen, allowing the experimenter to detect and record failures of the voice-activated relay. Immediately after each response, the prompt "Strategy?" appeared with the choices Transform, Count, Remember, and Other centered underneath. The participant reported their strategy choice and the experimenter recorded their answer using the number pad on the computer keyboard as a 1, 2, 3, or 4, respectively. The fixation dot then appeared and the experimenter recorded the answer to the problem using the number pad of the computer keyboard. Input of the answer cleared the screen, and the next trial commenced.

Results

Test-phase trials were classified into three categories based on operand familiarity. These categories included the practiced problems that were encountered during the practice phase, high-familiarity (HF) problems (both operands were familiarized during practice, but appeared in a new combination during test), and low-familiarity (LF) problems (neither operand was familiarized during the practice phase). Problem assignment to LF and HF conditions was counterbalanced. For participants who practiced problems based on 2, 5, 7, and 8, HF problems included two + seven, seven + two, two + eight, eight + two, five + seven, seven + five, five + eight, eight + five, and LF problems included three + six, six + three, three + nine, nine + three, four + six, six + four, four + nine, nine + four. For participants who practiced problems based on 3, 4, 6, 9, the HF problems included three + six, six + three, three + nine, nine + three, four + six, six + four, four + nine, nine + four, and LF problems included two + seven, seven + two, two + eight, eight + two, five + seven, seven + five, five + eight, eight + five. The remaining problems (one operand was familiarized and one was not) were considered fillers and were not analyzed because they could not be counterbalanced. The analysis also excluded ties (e.g., six + six) because their familiarity status is ambiguous. Additionally, we could not compare LF problems to practiced problems to measure effects of operand familiarity, because pair-specific facilitation from repeatedly solving the practiced problems would inevitably contaminate effects of operand familiarity. Thus, the critical comparisons for assessing familiarity compare the LF and the HF problems.

In addition to the familiarity factor, we calculated means separately for the first half of the test phase (Blocks 1 and 2) and the second half of the test phase (Blocks 3 and 4). Blocks 1 and 2 included the first encounter with each of the two orders of non-tie problems and Blocks 3 and 4 included the final encounters. It was important to divide the test phase this way because the repeated practice of each problem during the test phase will tend to equalize performance across the familiarity conditions. Consequently, we expected any differences across familiarity conditions to be clearest in the initial test blocks and to diminish over the final blocks.

We analyzed three dependent measures as a function of familiarity condition and block: percentage of retrieval reported, median response time (RT) for correct trials, and percentage of addition errors (i.e., incorrect answers). Less than 1% of RTs were spoiled by failures of the voice-activated switch. Means for retrieval percentage, RT, and error percentage appear in Table 2-1.

Table 2-1. *Mean Percentage of Reported Retrieval, Mean Response Time (RT) in Milliseconds, and Mean Percentage of Error as a Function of Block and Familiarity for Experiment 1*

	Blocks 1 and 2		*Blocks 3 and 4*	
Operand Familiarity	Mean	SD	Mean	SD
Retrieval				
Practiced	64.2	30.3	56.3	33.4
Low Familiarity (LF)	35.8	28.2	41.3	28.5
High Familiarity (HF)	40.4	29.0	40.0	29.5
RT				
Practiced	1206	324.8	1207	358.6
LF	1418	344.7	1281	330.9
HF	1452	353.2	1355	425.4
Error				
Practiced	5.8	11.6	5.4	10.4
LF	14.8	15.0	15.0	14.5
HF	12.9	12.4	14.0	15.6

Note. SD = sample standard deviation.

Examination of the table confirms that the practice phase was effective, in that repeatedly practicing problems promoted use of retrieval. Indeed, Table 2-1 shows that the mean percentage of reported retrieval for practiced problems was at least 23% higher than for any other condition in the first half of the test phase and at least 15% higher than any other condition in the second half of the test phase.[4] Table 2-1 shows that the mean RT for the practiced problems was substantially faster than the other conditions. Specifically, practiced problems were at least 200 ms (14%) faster than any other condition in the first half of the test phase and 74 ms (6%) faster in the second half of the test phase. Finally, Table 2-1 indicates that the practiced problems also had the lowest mean percentage of errors (5.8% vs. 13.9% on average for LF and HF problems in Blocks 1 and 2, and 5.4% vs. 14.5% in Blocks 3 and 4). These results confirm that the practice phase was effective in facilitating performance on the practiced operand pairs during the test phase. We next compare performance on LF and HF problems to determine if the practice phase also affected non-practiced problems.

Percent Retrieval for LF and HF Problems. A 2 × 2 ANOVA examined effects of familiarity (LF or HF) and block (Blocks 1 and 2 or Blocks 3 and 4). There was no main effect of block, $F(1, 59) = 2.03$, $MSE = 184.32$, $p = .159$, or familiarity, $F(1, 59) = .42$, $MSE = 397.07$, $p = .520$, but there was a Block × Familiarity interaction, $F(1, 59) = 4.21$, $MSE = 121.12$, $p = .045$. During Blocks 1 and 2, as predicted, the percentage of retrieval tended to be higher for problems with familiarized operands (40.4% for HF) than for problems with unfamiliarized operands (35.8% for LF), $t\ (59) = -1.64$, $SE = 2.80$, $p = .054$, one-tailed. The increase in retrieval for HF relative to LF problems was associated with a decrease in reported transformation (-4.0%) and "other" strategies (-0.83%), whereas reported counting did not differ between HF and LF problems (+0.21%). During Blocks 3 and 4, however, the trend towards greater retrieval use disappeared (40.0% for HF vs. 41.3% for LF), $t\ (59) = .407$, $SE = 3.07$, $p = .685$, two-tailed. The significant Familiarity × Block interaction shows that, early in the test phase, the percentage of retrieval reported tended to be higher for problems composed of familiarized operands than problems with unfamiliar operands. The strategy differences disappeared over the final blocks, presumably because practicing all of the problems during Blocks 1 and 2 tended to equalize performance for HF and LF problems during Blocks 3 and 4.

RT for LF and HF Problems. A 2 × 2 ANOVA of the means in Table 2-1 confirmed the standard effect of practice, with mean latency decreasing from Blocks 1 and 2 (1359 ms) to

---

[4] The rate of retrieval observed here for the practiced problems is similar to previous studies examining addition retrieval using word format (Campbell et al., 2004; Campbell & Penner-Wilger, 2006).

Blocks 3 and 4 (1281 ms), $F(1, 59) = 14.69$, $MSE = 55417.06$, $p < .001$. The main effect of familiarity approached significance, $F(1, 59) = 2.81$, $MSE = 62586.54$, $p = .099$: mean RT for HF problems was 1404 ms compared to 1350 ms for LF problems. There was no evidence for a Block × Familiarity interaction, $F(1, 59) = 0.55$, $MSE = 44248.25$, $p = .460$. The tendency for slower mean RTs for HF compared to LF problems is somewhat surprising given the evidence that retrieval usage tended to be higher for HF problems, at least during test Blocks 1 and 2. Generally, retrieval is faster than procedural strategies; therefore, increased retrieval ought to be associated with faster RTs on average.

Although RT tended to be slower overall for HF relative to LF problems, we might still expect familiarity to have consistent effects on both RT and retrieval usage across participants. In fact, the Pearson correlation relating the difference in mean RT of all trials (HF minus LF) and mean difference in the percentage of reported retrieval (HF minus LF) was $r(58) = -.32$, $p = .011$. This correlation indicates that increased retrieval usage for HF relative to LF problems was associated with faster addition RTs, presumably, because retrieval is generally faster than arithmetic procedures. Consistent with this interpretation, the correlation disappeared when the analysis included only the RT scores of solutions that were retrieved (i.e., procedure solutions were excluded), $r(39) = .12$, $p = .46$.

Error Rates for LF and HF Problems. The mean percentage of errors overall for LF and HF problems was 14.2%. The Block × Familiarity, 2 × 2 ANOVA of the means in Table 2-1 indicated no significant effects, all *F*s < 1. Unlike the RT analysis, there was no evidence that error differences (HF problems minus LF problems) were predicted by mean differences in reported retrieval (HF minus LF), $r(58) = -.15$, $p = .244$.

Discussion

Experiment 1 provided evidence that familiarizing participants with specific operands increased their reported memory retrieval for unpractised problems with highly familiar operands (HF) compared to those composed of less familiar operands (LF). Furthermore, across the group of 60 participants, the difference between retrieval rate for HF and LF problems was negatively correlated with the corresponding difference in mean RT between HF and LF problems. This correlation indicates that, across individuals in the group, greater use of retrieval for HF compared to LF problems was associated with faster RTs for HF compared to LF problems. We would expect this relation because, on average, retrieval is faster than a nonretrieval strategy; consequently, more retrieval ought to be associated with faster RT. Nonetheless, overall, mean RT for HF problems tended to be longer than for LF problems.

One possible explanation for slower retrieval RTs for HF relative to LF problems is that practicing a subset of problems (e.g., five + two and seven + eight) produces retrieval-induced forgetting for other problems composed of those operands (e.g., five + seven and two + eight). Retrieval-induced forgetting is the phenomena whereby increased memory strength of one item through practice reduces the accessibility of related items, either because of interference or inhibition (Anderson & Bell, 2001; Phenix & Campbell, 2004). An analysis of RTs that included only trials on which retrieval was reported confirmed that retrieval, per se, was slower for HF (1282 ms) relative to LF (1190 ms) problems, $F(1, 40) = 4.50$, $MSE = 76224.98$, $p = .040$.[5] Retrieval-induced forgetting of HF compared to LF items would tend to produce slower RTs for HF problems and potentially work against the facilitating effects of operand familiarity. Experiment 2 was designed to eliminate these potentially opposing influences. Specifically, Experiment 2 was identical to Experiment 1 except that, rather than familiarizing operands through addition (two + five = ?), participants repeatedly performed magnitude comparisons on the corresponding operand pairs (e.g., two   five, which is larger?). The comparison task should familiarize participants with the operand pair, but should eliminate the strengthening of competing addition associations that potentially promoted retrieval-induced forgetting in Experiment 1.

Experiment 2

Method

Participants. Thirty-eight introductory Psychology students (29 female, 9 male) participated for course credit at the University of Saskatchewan. The mean age of participants

---

[5] The retrieval-only analysis included the 41 participants who had at least one valid retrieval RT in each of the four Familiarity × Block cells.

was 20.3 years; 34 were right handed, and 31 claimed English as their first language for arithmetic.

Stimuli, Design, and Procedure. With the exception of the practice phase, Experiment 2 used the same method as Experiment 1. In Experiment 1, practice consisted of a subset of possible addition problems from one of two counterbalanced families of operands (e.g., 2, 5, 7, 8, or 3, 4, 6, 9). In Experiment 2, we replaced addition problems in the practice phase with magnitude comparison problems. The magnitude comparison task required participants to report the larger of the two numbers provided (e.g., "two five" respond 5). For tie problems (e.g., "two two") participants were instructed to state the tie number. Five spaces separated the numbers, with the stimuli centered on the middle separating space. Problem subsets were the same as those utilized in the practice phase of Experiment 1 (two five, five two, seven eight, eight seven, two two, five five, seven seven, eight eight, or three four, four three, six nine, nine six, three three, four four, six six, nine nine). Counterbalancing of the practice sets was done by participant number (even = 2, 5, 7, 8; odd = 3, 4, 6, 9). The eight practice problems were randomly repeated eight times for a total of 64 practice trials. After completion of the practice phase, participants proceeded with the test phase, which was comprised of four blocks of 36 addition problems as in Experiment 1. Instructions and procedures for the test phase, including strategy reports, were duplicates of those used in Experiment 1.

Results

Less than 2% of RTs were spoiled by failures of the voice-activated switch. Means for retrieval percentage, RT, and error percentage appear in Table 2-2 as a function of problem type (practiced, LF, and HF) and block (Blocks 1 and 2 vs. Blocks 3 and 4). Unlike Experiment 1, we did not expect pair-specific practice effects to transfer from the comparison task to the addition task. Consequently, in Experiment 2, a comparison of practiced problems to LF problems might provide a sensitive measure of the effects of operand familiarity. When we matched practiced operands to LF counterparts, however, there was evidence of pair-specific interference with a higher addition error rate for practiced pairs (11.5%) compared to LF problems (5.3%), $F(1, 37)$

= 6.11, *MSE* = 242.82, *p* = .018.[6] Given this evidence of pair-specific interference with practiced problems, we cannot confidently compare LF problems to practiced problems to measure effects of operand familiarity. Thus, as in Experiment 1, the critical comparisons for assessing familiarity effects compared the LF to the HF problems.

Table 2-2. *Mean Percentage of Reported Retrieval, Mean Response Time (RT) in Milliseconds, and Mean Percentage of Error as a Function of Block and Familiarity for Experiment 2*

	Blocks 1 and 2		*Blocks 3 and 4*	
Operand Familiarity	Mean	SD	Mean	SD
Retrieval				
Practiced	52.0	29.9	55.9	33.1
Low Familiarity (LF)	48.0	28.7	52.6	29.9
High Familiarity (HF)	56.6	29.0	54.9	30.4
RT				
Practiced	1319	363.7	1183	414.0
LF	1272	292.2	1136	257.0
HF	1280	260.6	1186	305.5
Error				
Practiced	11.2	16.1	11.8	18.1
LF	12.5	14.2	14.8	13.6
HF	7.9	10.2	11.5	10.6

Note. SD = sample standard deviation.

Percent Retrieval for LF and HF Problems. A 2 × 2 ANOVA examined effects of familiarity (LF or HF) and block (Blocks 1 and 2 or Blocks 3 and 4). There was no evidence that reported use of retrieval differed between early and later test blocks, *F*(1, 37) = 0.61, *MSE* =

[6] For participants who practiced comparison pairs corresponding to the addition problems 2 + 5, 5 + 2, 7 + 8, and 8 + 7, the LF counterparts were 3 + 4, 4 + 3, 6 + 9, and 9 + 6, whereas for participants who practiced pairs corresponding to 3 + 4, 4 + 3, 6 + 9, and 9 + 6, the LF counterparts included 2 + 5, 5 + 2, 7 + 8, and 8 + 7.

136.05, p = .439, but participants reported more retrieval for HF (55.8%) than LF problems (50.3%), $F(1, 37)$ = 5.35, *MSE* = 209.40, p = .026. There was no Block × Familiarity interaction, $F(1, 37)$ = 1.85, *MSE* = 200.06, p = .181, although the trend was similar to Experiment 1. During Blocks 1 and 2, the percentage of retrieval was higher for problems with familiarized operands (56.6% for HF) than for problems with unfamiliarized operands (48.0% for LF), and this difference was statistically robust, t (37) = -2.59, *SE* = 3.30, p = .007, one-tailed. As in Experiment 1, the increase in retrieval for HF relative to LF problems was primarily associated with a decrease in reported transformation (-4.6%) and "other" strategies (-2.6%), whereas reported counting did not differ between HF and LF problems (-1.3%). During Blocks 3 and 4, however, the rate of retrieval was similar for HF and LF problems (54.9% for HF vs. 52.6% for LF), t (37) = -0.71, *SE* = 3.26, p = .243, one-tailed. Thus, given the results for Blocks 1 and 2, Experiment 2 confirmed the finding of Experiment 1 that operand familiarity increased participants' reported use of direct memory retrieval.

RT for LF and HF Problems. A 2 × 2 ANOVA of the means in Table 2-2 confirmed the standard effect of practice, with mean latency decreasing from Blocks 1 and 2 (1276 ms) to Blocks 3 and 4 (1161 ms), $F(1, 37)$ = 27.68, *MSE* = 18104.51, p < .001. There was no evidence for a main effect of familiarity, $F(1, 37)$ = 0.63, *MSE* = 49761.67, p = .431, or a Block × Familiarity interaction, $F(1, 37)$ = 0.87, *MSE* = 19549.86, p = .356. Unlike Experiment 1, there was no evidence that the differences in mean RT between problem sets (HF problems minus LF problems) was correlated with the mean difference in reported retrieval (HF minus LF).

Error Rates for LF and HF Problems. There was an increase in error rates from Blocks 1 and 2 (10.2%) to Blocks 3 and 4 (13.2%), $F(1, 37)$ = 5.69, *MSE* = 58.57, p = .022.[7] The error rate showed a trend to be lower for HF (9.7%) than for LF (13.7%) problems, $F(1, 37)$ = 3.77, *MSE* = 157.14, p = .060. There was no evidence of a Block × Familiarity interaction, $F(1, 37)$ = 0.19, *MSE* = 88.24, p = .668. In Experiment 1, we found that the difference in retrieval usage between HF and LF problems predicted the RT difference between HF and LF problems, reflecting the fact that retrieval tends to be more efficient than procedures. In Experiment 2, this

[7] The increase in errors over blocks in Experiment 2 might simply reflect a speed-accuracy trade-off, as RTs decrease substantially. This trend was also apparent, but not statistically significant, in Experiment 1.

relationship emerged in the analysis of errors rather than in RT. Specifically, the Pearson correlation relating mean difference in error rate (HF problems minus LF problems) and mean difference in the percentage of reported retrieval (HF minus LF) was $r(36) = -.59$, $p < .001$. This indicates that increased retrieval usage for HF relative to LF problems was associated with fewer addition errors, presumably because retrieval is generally more accurate than procedures. Consistent with this interpretation, the correlation was substantially weakened when the analysis included only retrieval trials (i.e., procedure trials were excluded), $r(30) = -.28$, $p = .13$.

Discussion

Experiment 2 provided further evidence for increased memory retrieval for problems composed of familiarized operands (HF) compared to less familiar operands (LF). The HF problems also were associated with a trend towards fewer errors than LF problems. Furthermore, differences in error rates between HF and LF problems correlated negatively with the corresponding differences between the percentage of reported retrieval for HF and LF problems. This indicates that, across participants, greater use of retrieval for HF relative to LF problems was associated with fewer errors for HF compared to LF problems. Memory retrieval for simple addition tends to be more accurate than non-retrieval strategies (Campbell & Xue, 2001). Consequently, increased retrieval for HF relative to LF problems would also tend to reduce errors for HF problems.

General Discussion

The purpose of our experiments was to determine if operand familiarity affected the probability of memory retrieval for simple addition. In Experiment 1, the practice phase involved a subset of the simple addition problems, whereas the practice phase in Experiment 2 involved magnitude comparisons. For the test phase in both experiments, participants solved addition problems comprised of familiarized or unfamiliarized operands. In Experiment 1 the reported rate of direct memory retrieval was 13% higher for HF problems relative to LF problems in Blocks 1 and 2. In Experiment 2, the reported rate of retrieval was 18% higher for HF compared to LF problems in Blocks 1 and 2. In both experiments, individual differences in the higher rate of retrieval of HF relative to LF problems were correlated with corresponding differences in performance. In Experiment 1, increased retrieval usage for HF relative to LF problems was associated with faster addition RTs (r = -.32), whereas in Experiment 2 it was associated with reduced errors (r = -.59). These correlations indicate that the greater use of retrieval for HF compared to LF problems was associated with better performance on HF problems compared to LF problems.

Our results suggest that the cue familiarity effects observed in a variety of non-arithmetic memory tasks are also observed in memory for elementary arithmetic facts. In our experiments, we used written word operands because the reported rate of direct memory for addition facts with word operands is usually quite low. This provided the opportunity to observe increased retrieval usage in connection with manipulations of operand familiarity. According to Koriat and Levy-Sadot (2001), cue familiarity stimulates more memory search. Thus, when the familiarity of word format problems was increased through pre-exposure during the practice phase, the cue-familiarity of these problems was increased, promoting greater use of retrieval.

It is important to emphasize that the effects of familiarization on retrieval usage did not lead participants to use retrieval inappropriately. As we have already pointed out, increased use of retrieval for HF relative to LF problems was associated with better performance for HF relative to LF problems, in terms of either speed (Experiment 1) or accuracy (Experiment 2). Apparently, the sense of familiarity generated by HF problems allowed participants to calibrate their strategy choices to use retrieval more effectively for HF compared to LF problems. Given that the sense of familiarity attributed to HF problems would actually reflect a source confusion (i.e., the HF operands were familiarized in the context of other problems), how would such calibration work? Why would the "false familiarity" presented by HF problems not lead to an inappropriate use of retrieval? It is possible that participants in our experiment globally underestimated the potential utility of retrieval for simple addition problems, possibly because our instructions raised their awareness of procedural strategies (Kirk & Ashcraft, 2001). A tendency to underestimate the utility of direct retrieval would also be promoted by the written word format we used (four + eight), which discourages retrieval (Campbell et al., 2004). In this context, participants might set an unnecessarily high criterion for the level of familiarity that a problem must produce before retrieval is attempted. After HF operands were familiarized in the practice phase, the sense of familiarity generated by HF problems during the test phase was more likely to exceed the criterion for retrieval than that generated by LF problems.

The increase in retrieval reported for HF relative to LF problems was quite small in absolute terms (4.6% in Experiment 1 and 8.6% in Experiment 2). By contrast, operand format (e.g., 4 + 8 vs. four + eight) produces absolute differences in retrieval usage for simple addition of about 15% favouring the Arabic format (70%) relative to word format (55%) (Campbell et al., 2004). The effect of problem size on reported use of retrieval for simple addition is even larger (e.g., 49% with sum $N + M \leq 10$, 77% for $N + M > 10$) (Campbell et al., 2004; Campbell & Penner-Wilger, 2006; LeFevre et al. 1996b). Thus, based on the present experiment, the effects of familiarity on strategy choice for simple addition are relatively modest compared to some other known influences. One implication of this is that it is unlikely that the effects of format and problem size on retrieval usage simply are manifestations of familiarity, because the effects of familiarity on strategy choice, at least as measured in these experiments, are much smaller than the effects of format or problem size.

Nonetheless, operand familiarity is potentially a component of problem size and format effects on addition strategy choice. Campbell et al. (2004) speculated that lower rates of retrieval for addition in written word compared to digit format might reflect lower familiarity of arithmetic problems in the word format. The present results are consistent with the possibility that familiarity of the written word operands does contribute to this effect. Similarly, small number problems are encountered more often than large number problems (Zbrodoff & Logan, 2005) and thus, operand or problem familiarity might contribute to participants selecting retrieval more often for small than for large problems.

Although the effects of our familiarity manipulation on strategy choice were relatively small, they were nonetheless substantial, particularly in Experiment 2. These effects are not predicted by any current model of strategy choice for simple arithmetic (e.g., Rickard, 2005; Siegler & Lemaire, 1997; Shrager & Siegler, 1998). The current models emphasize the relative strength of the correct-answer association as the major factor in the use of retrieval vs. use of a procedural strategy. They do not incorporate a role for problem familiarity independent of associative strength. Our results indicate that models of strategy choice for arithmetic will need to incorporate mechanisms of familiarity and source confusion such as those found in the strategy choice models of Reder (cf. Schunn et al., 1997) or Koriat (Koriat & Levy-Sadot, 2001).

Another important contribution of our experiments is the demonstration that effects of familiarity on addition strategy choice are not operation specific. We observed increased retrieval for HF relative to LF problems both when operands were familiarized in the context of addition (Experiment 1) and when they were familiarized in the context of number comparison (Experiment 2). Is it possible, however, that participants in the number comparison task performed addition, although they were not requested to do so? This seems unlikely because there was no evidence of addition facilitation for problems corresponding to the practiced comparison pairs (see Table 2-2). On the other hand, addition performance might incorporate an operand comparison stage (Butterworth, Zorzi, Girelli & Jonckheere, 2001). In this case, the addition and comparison tasks in the present experiment would include common processing stages. This overlap might contribute to familiarity acquired in the comparison task transferring to the addition task.

Nonetheless, the effects of familiarity on strategy choice for addition did not depend on the component operands being familiarized in the context of addition. It might be sufficient to familiarize the operands in a task that only requires the processing of the numerical meaning of the operands. In fact, it is possible that mere exposure to single digit numbers one by one might be sufficient. Based on our experiments, we cannot say what critical elements are required for familiarity to influence addition strategy choice. The number-word operands used for our comparison task were identical in shape, size, and location to those used for addition. These superficial similarities might be important. Similarly, the comparison task required an intentional semantic analysis of numbers, but we do not know if this is crucial to influence subsequent addition strategy choice. Finally, we do not know the duration of the familiarity effects generated in our experiments, but Schunn et al. (1997) observed effects of familiarity on intention to retrieve after a 24-hour interval. Thus, operand familiarization might have long lasting effects on retrieval.

Conclusions

Two experiments investigated the role of operand familiarity in adults' performance of simple addition. We manipulated the familiarity of a subset of operands by having adults repeatedly practice specific additions (two + five = ?; Experiment 1) or magnitude comparisons (two five, choose the larger; Experiment 2). Both experiments provided evidence that pre-exposure to single-digit operands increased reported use of direct retrieval for new combinations of the familiarized operands. RT and error patterns across experiments also supported the conclusion that increased use of retrieval facilitated performance. The results extend the domain of application of cue familiarity effects (cf. Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992; Schunn et al., 1997) to include elementary arithmetic. Accordingly, models of strategy choice for arithmetic should consider the potentially important role of operand familiarity.

CHAPTER 3

A majority of the research on basic arithmetic and basic arithmetic strategy use has concentrated on problems presented in Arabic digit format. After Arabic digit format, the next runners up are studies that contrast performance on Arabic digit format with problems in written word format. This class of studies has consistently shown processing differences that are format dependant and these differences have been marshalled as evidence in the debate over the organization of arithmetic solution architecture in the human mind. On the other hand, problems presented as auditory-verbal stimuli had only appeared in a single study in the area at the time this research was undertaken. This state of affairs was especially surprising when two major points are considered in tandem: 1) that verbal arithmetic is arguably as common as Arabic digit format, if not more so, as a format for early-life practice and everyday basic arithmetic, and 2) that if written word arithmetic afforded insights into the cognitive architecture of solution processes, it should be at least equally likely that the more commonly used verbal-auditory format would as well.

LeFevre, Lei, Smith-Chant, and Mullins (2001) examined effects of auditory vs. Arabic visual presentation formats on performance of simple multiplication. They observed a smaller problem-size effect (RT increases with numerical size) with auditory compared to Arabic stimuli. If this arises during problem encoding, as opposed to during subsequent calculation processes, we would expect comparable Format × Problem Size interactions for both multiplication and addition. For multiplication, we replicated the finding of a smaller problem-size effect for auditory compared to Arabic format, but found the opposite pattern for addition whereby the problem-size effect was larger with auditory than Arabic stimuli. Decomposition of mean RT into its ex-Gaussian components, μ and τ, demonstrated that the triple-interaction arose entirely in connection with τ. This suggests that the effects of auditory vs. Arabic format on RT substantially reflected format-related shifts in the use of procedural strategies.

This chapter has been previously published in a journal published by the American Psychological Association:

Metcalfe, A. W. S., & Campbell, J. I. D. (2008). Spoken numbers vs. Arabic numerals: Differential effects on adults' multiplication and addition. *Canadian Journal of Experimental Psychology*, *62*, 56-61.

Auditory Modality and Arithmetic Performance

LeFevre, Lei, Smith-Chant and Mullins (2001) compared English and Chinese speakers' performance on simple multiplication problems (e.g., 2 × 3, 9 × 6) presented in auditory and Arabic format. Of particular interest here was the observation that, for the English-speaking group, the *problem-size effect* (PSE) was 25% larger with Arabic than with auditory presentation format (see LeFevre, et al., 2001, Table 1, p. 280). The PSE is the ubiquitous finding that response time (RT) tends to be slower for large (e.g., 9 × 6) relative to small problems (e.g., 2 × 3; see Zbrodoff & Logan, 2005, for a review of the PSE). LeFevre, et al.'s observation of a smaller PSE for auditory verbal format ("four times eight") compared with Arabic format (4 × 8) is interesting because it contrasts with previous evidence of a larger PSE in latencies for written verbal format (four × eight) compared to Arabic (4 × 8; e.g., Campbell, 1994; Noël, Fias, & Brysbaert, 1997). The source of the Format × Problem Size interactions with written number words vs. Arabic format has been controversial. Some researchers have argued that the effect arises at encoding (McCloskey, Macaruso, & Whetstone, 1992; Noël et al., 1997) whereas others have argued that it occurs during calculation (Campbell, 1994; Campbell & Epp, 2005). Similarly, we consider if the Format × Problem Size pattern observed by LeFevre et al. with auditory and Arabic formats arises during encoding or at another stage of processing.

If LeFevre et al.'s (2001) Format × Problem Size pattern was the result of encoding differences for auditory and Arabic operands, then we would expect the same form of the interaction to be present in both simple multiplication and addition. Fact retrieval for both operations is governed by the same representational principles (Campbell, Fuchs-Lacelle, & Phenix, 2006), and it is a common assumption that both operands are converted to an internal quantity representation before solution processes proceed (Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Campbell, 1994; Verguts & Fias, 2005). Thus, multiplication and addition are expected to engage similar encoding processes. Therefore, if we find different forms of the Format × Problem Size interaction for multiplication and addition, this would suggest that format affects operation-specific processes that occur after encoding.

Method

Participants

Seventy-four University of Saskatchewan students (39 men and 35 women) between the ages of 17 and 41 years (*M* = 24) received $5 for participation in the experiment. All participants reported normal or corrected-to-normal vision.

Stimuli and Apparatus

Participants wore headphones, and a lapel microphone was used to control a software clock accurate to ± 1 ms. The microphone was connected to the computer via a relay switch. Stimuli were the simple multiplication and addition problems composed of the numbers 2 through 9. The problems were presented in visual form as Arabic digits or presented auditorily as English words. Arabic problems were displayed sequentially in horizontal orientation with left and right operands separated by the operation sign (+ or ×) and a space on either side of the operation sign. The three components of the problem—left operand, operation sign, and right operand—were presented serially for 250 ms each. Problems appeared on a computer monitor as white characters approximately 6 mm high × 4 mm wide against a black background.

English names for the numbers 2 through 9 and the words "times" and "plus" were digitally recorded in an adult male voice. The auditory files for the number words and signs were edited to a constant length of approximately 250 ms. The auditory files were used to construct a set of auditory arithmetic problems that paralleled the Arabic problems.

Design

There are 36 different addition and multiplication problems involving the operands 2 through 9 when operand order is ignored (i.e., when 2 + 5 and 5 + 2 are treated as the same problem). Each participant received four blocks of 72 trials (i.e., arithmetic problems). Problems alternated between auditory and Arabic presentation. Even-numbered participants began with a multiplication block and odd-numbered participants began with an addition block. Blocks alternated between multiplication and addition. Within each block, the order of problems was randomized with the constraint that auditory and Arabic format versions of the same problem (i.e., 2 + 5, "two plus five") were always separated by at least 20 trials. Operand order was counterbalanced. In the first block, approximately half of the nontie problems (i.e., problems consisting of different operands such as 4 + 8) were randomly selected to be tested with the smaller operand on the left (Arabic presentation) or spoken first (auditory presentation). On the basis of this selection procedure result, operand order then alternated across blocks.

Procedure

Testing occurred in a quiet, dimly lit room with an experimenter present. The following instructions were presented to participants:

> There will be 4 blocks of simple addition and multiplication. In each block there will be 72 problems in a random order. We are interested in speed of memory retrieval for arithmetic facts, so your task is to state the answer as quickly and as accurately as possible. Presentation modality will alternate between audition and vision. That is, you will hear the problems spoken through the headphones on one trial, then, on the next trial, a different problem will be presented visually. To make the timing of auditory and visual displays comparable, the two digits in the visually presented problems will appear about 1/2 second apart with the left digit appearing first. Occasional errors are normal and should not concern you. A warning dot will flash twice before each problem appears. Please always try to respond as quickly as possible.

The participant and experimenter viewed separate monitors with the participant facing the monitor at an approximate distance of 60 cm. Blocks of trials were separated by approximately 15 s and were initiated by the experimenter. Each block was preceded by the appearance of either the word MULTIPLICATION or ADDITION whichever was appropriate to the block. Before each trial a prompt for the modality of the upcoming problem flashed twice for a total of 250 ms in the center of the screen, as a yellow V (for visual) or a white A (for auditory). After the 250-ms format prompt, a fixation dot appeared at the center of the screen, and flashed twice over a period of 1 s. Serial presentation of the problem components (i.e., first operand, operand sign, and second operand) commenced on what would have been the third flash of the fixation dot. For the Arabic stimuli, the operation sign appeared at the fixation point. Response timing began at the offset of the second operand and stopped when the microphone detected the verbal response. When the microphone detected a signal, the fixation dot appeared immediately, which allowed the experimenter to detect microphone failure. During this postresponse interval the experimenter recorded the response to the problem using the keyboard's number pad and the fixation dot remained until the response was recorded to a minimum of 2.5 s. If the response was recorded within 2.5 s the inter-trial interval was 4.5 s. After the experimenter recorded the response to the trial, the fixation was removed and the next trial commenced with presentation of the modality prompt (yellow V or white A). No feedback regarding accuracy or response time was given to the participants. Trials where the microphone failed to detect the participant's response or responded to extraneous noise were marked as spoiled trials. Approximately 60 min was required to test each participant.

Results

To operationalize problem size for both addition and multiplication, small problems were defined as those with operand pairs whose product was less than or equal to 25; otherwise, the problem was large (Campbell & Xue, 2001; Campbell, Parker, & Doetzel, 2004).

Response Time

A total of 644 RTs (4.2%) were discarded as outliers greater than 3 standard deviations from each Operation × Format cell. About 2% of RTs were spoiled because of failures of the voice key. Table 3-1 presents mean response times for correct trials for small and large multiplication and addition problems presented in auditory and Arabic format.[8] A 2 × 2 × 2

---

[8] LeFevre et al. (2001) measured RT from the onset of the second operand, whereas we commenced measurement at the termination of the second operand, which was always 250 ms in duration. This would largely account for the 265 ms difference in mean RT observed here (546 ms) compared to LeFevre et al. (811 ms).

repeated measures analysis of variance (ANOVA) was performed using the factors operation (multiplication and addition), format (auditory and Arabic), and problem size (small and large). For operation, addition was faster (521 ms) than multiplication (566 ms), $F(1, 73) = 11.53$, $MSE = 25645.76$, $p = .001$. Auditorily presented problems were solved more quickly (536 ms) than problems in digit format (552), $F(1, 73) = 5.14$, $MSE = 7361.86$, $p = .026$. Small problems were solved more quickly (486 ms) than large problems (601 ms), $F(1, 73) = 147.40$, $MSE = 13293.42$, $p < .001$, and, as is usually found (Campbell & Xue, 2001), the PSE was larger for multiplication (+132) than addition (+98 ms), $F(1, 73) = 8.34$, $MSE = 5151.98$, $p = .005$. Most important, there was a robust three way interaction of problem size, operation, and format, $F(1, 73) = 10.70$, $MSE = 2284.72$, $p = .002$.

Table 3-1. *Overall μ and τ Solution Response Times (RT) in Milliseconds and Mean Percent Error for Multiplication and Addition as a Function of Format and Problem Size*

	Multiplication			
Problem Size	Mean (SE)	μ (SE)	τ (SE)	Error
Auditory				
Small (S)	498 (17)	337 (9)	163 (12)	4.5
Large (L)	616 (23)	428 (16)	194 (17)	16.8
L – S	118	91	31	12.3
Arabic				
Small	502 (13)	384 (7)	119 (10)	4.1
Large	649 (24)	455 (13)	204 (19)	16.0
L – S	147	71	85	11.9
	Addition			
Auditory				
Small	460 (17)	326 (10)	136 (11)	4.5
Large	569 (23)	406 (16)	166 (13)	8.6
L - S	109	80	30	4.1
Arabic				
Small	485 (12)	363 (8)	125 (11)	4.8
Large	571 (21)	442 (16)	132 (11)	7.7
L - S	86	79	7	2.9

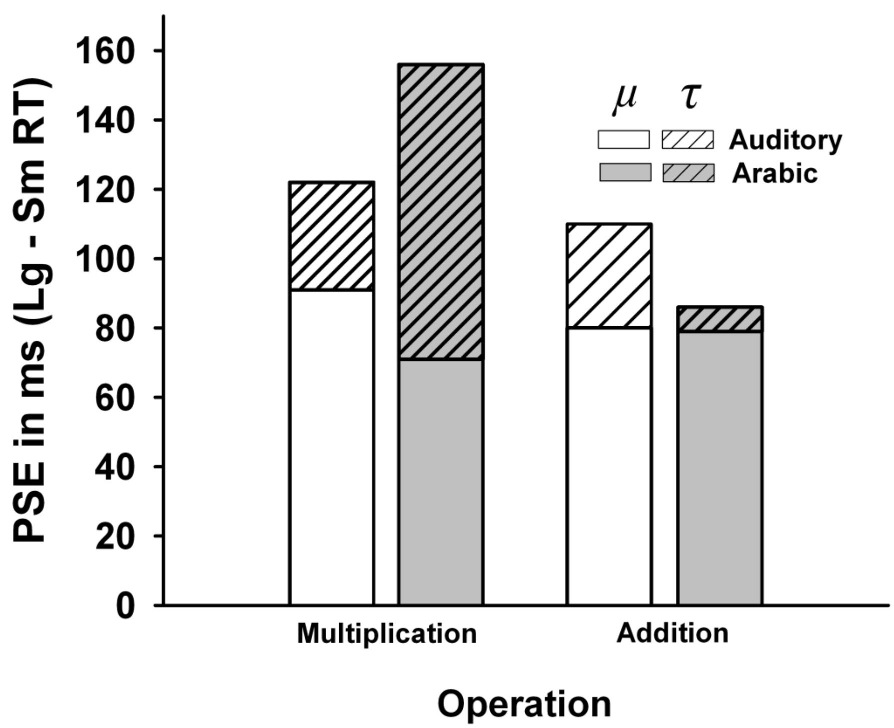
Note. SE = standard error of the mean.

Separate 2 × 2 ANOVAs with factors of format (auditory and Arabic) and problem size (small and large) were performed for the multiplication and addition tasks. For multiplication, the PSE was smaller with auditory format (118 ms) than Arabic format (147 ms), $F(1, 73) = 6.88$, $MSE = 2224.57$, $p = .011$, which replicates the pattern observed by LeFevre et al. (2001). In contrast, for addition, the PSE was larger with auditory format (109 ms) compared to Arabic format (86 ms), $F(1, 73) = 5.03$, $MSE = 1887.19$, $p = .028$. Thus, the three-way Operation × Format × Problem Size interaction occurred because the pattern of the Format × Problem Size interaction was in opposite directions for multiplication and addition.

We pursued the three-way interaction by analyzing ex-Gaussian components of the latency distributions. The ex-Gaussian model includes a normally distributed component corresponding to the main body of the RT distribution and an exponentially distributed upper tail corresponding to relatively slow RTs contributing to skew. The parameter μ estimates the mean of the normal component and τ estimates the mean of the exponential tail. Overall mean RT is simply the sum of μ plus τ. Analysis of μ and τ can be informative about the solution strategies used for simple arithmetic; particularly to discriminate the contribution of direct memory retrieval from elaboration procedures such as counting, repeated addition, and so forth. Previous research has demonstrated that use of procedural strategies (which are generally slower than direct retrieval and therefore more likely to fall in the upper tail of the RT distribution) are expressed more in τ than in μ (Campbell & Penner-Wilger, 2006; Penner-Wilger, Leth-Steensen & LeFevre, 2002). For example, Campbell and Penner-Wilger analyzed μ and τ for adults' simple addition and recorded self-reported strategy (i.e., retrieval or procedure) trial by trial. They isolated the contribution of procedure trials by contrasting all-trials RTs with retrieval-trials RTs. The results confirmed that procedure trials contributed to mean RT primarily by influencing τ rather than μ.

We decomposed mean RT into its ex-Gaussian components μ and τ using QMPE v2.18 for Windows (Cousineau, Brown, & Heathcote, 2004; Heathcote, Brown, & Mewhort, 2002). Our goal was to determine if the three-way interaction in the analysis of mean RT was associated exclusively with τ or with μ. Table 3-1 includes the estimated means for μ and τ for each cell of the Operation × Format × Problem Size analysis.[9] Figure 3-1 depicts the PSEs (i.e., L-S in Table 3-1) associated with μ and τ as a function of format and operation. Figure 3-1 depicts the PSEs (i.e., large-small in Table 3-1) associated with μ and τ as a function of format and operation. Figure 3-1 shows that differences in mean PSE across formats were tied strongly with τ but not with μ. ANOVAs confirmed a robust three-way interaction on τ, $F(1, 73) = 10.70$, $MSE = 5033.68$, $p = .002$, but there was no evidence of a three-way effect on μ, $F(1, 73) = 1.093$, $MSE = 3003.77$, $p = .299$. As we explain in the Discussion section, this raises the possibility that the three-way interaction in mean RT substantially reflects format-related strategy shifts.

Figure 3-1. Ex-Gaussian μ and τ Response Time Differences in Answer Latency Computed Between Large and Small Problems as a Function of Auditory and Arabic Format for Multiplication and Addition



There were other revealing effects of format in the analyses of μ and τ. In the analysis of μ, there was a robust main effect of format, $F(1, 73) = 39.153$, $MSE = 9088.017$, $p < .001$, with

[9] For relatively small sample sizes the values of μ and τ estimated by the QMPE program do not necessarily sum exactly to the mean RT.

mean μ 36 ms faster for auditory compared to Arabic stimuli. For τ, however, the pattern reversed and presented a robust main effect with mean τ 20 ms faster for Arabic compared with auditory stimuli, $F(1, 73) = 8.244$, $MSE = 7128.307$, $p = .005$. Before discussing implications of these findings, we will first examine errors.

Errors

Table 3-1 presents mean percentage of errors for small and large addition and multiplication problems presented in auditory and Arabic format. A 2 × 2 × 2 ANOVA with factors of operation, format, and problem size was conducted. For main effects of operation and problem size, there was the usual pattern of more errors for multiplication problems (10.4 %) than addition problems (6.4 %), $F(1, 73) = 21.03$, $MSE = 111.57$, $p < .001$, and more errors on large problems (12.3 %) than small problems (4.5 %), $F(1, 73) = 95.99$, $MSE = 93.67$, $p < .001$. These main effects were qualified by an Operation × Problem Size interaction whereby the PSE on errors was larger for multiplication problems (+12.1%) than addition problems (+3.5 %), $F(1, 73) = 48.02$, $MSE = 56.86$, $p < .001$. There was no evidence that format affected overall error rate (all *p*s > .10).

To pursue the possibility of more subtle effects of format on errors, we classified errors into to several subclasses. Table 3-2 presents the subclass definitions and frequencies as a function of operation, format, and problem size (see Campbell, 1994, for a detailed discussion of these error subclasses). ANOVAs performed on each subclass revealed no main or interaction effects involving the format factor. In contrast, LeFevre et al. (2001) found that multiplication errors with auditory presentations were more likely to contain operand intrusions than with Arabic presentation. Operand intrusions are errors that contain one of the operands (e.g., 9 × 6 = 36, 6 × 9 = 63, 8 × 4 = 24, 4 × 8 = 28). Here, the overall proportion of multiplication errors involving intrusions was 35% with Arabic stimuli and 37% with auditory stimuli, and the frequencies (146 and 162 for Arabic and auditory, respectively) were statistically equivalent, $F(1, 73) = 0.86$, $MSE = .50$, $p = .358$. A common feature of intrusion errors is that the position of the intruder in the problem (i.e., the first or second operand) is often congruent with the position of the intrusion in the stated answer (decade or unit) (e.g., 9 × 6 = 36 is congruent whereas 6 × 9 = 36 is incongruent; Campbell, 1994; LeFevre et al., 2001; Noël et al., 1997). Congruent intrusions were more frequent than incongruent intrusions, $F(1, 73) = 15.8$, $MSE = .68$, $p < .001$, but this did not differ between Arabic (95 vs. 51) and auditory format (99 vs. 63), $F(1, 73) = 0.31$, $MSE = .35$, $p = .851$. Thus, we did not find evidence that the rate of operand intrusions differed between the Arabic and auditory formats. We turn to implications of this in the following discussion.

Table 3-2. *Total number of Errors as a Function of Type, Operation, Format, and Problem Size*

	Multiplication					
	Tabled Errors					
Problem Size	Congruent	Incongruent	Non-Intrusions	Operation	Misc.	Total
Auditory						
Small	25	14	39	8	8	94
Large	74	49	168	3	54	348
Arabic						
Small	13	13	37	16	6	85
Large	82	38	165	4	43	332
	Addition					
Auditory						
Small	7	1	54	30	1	93
Large	11	11	153	2	1	178
Arabic						
Small	8	5	52	34	0	99
Large	15	8	134	1	1	159

Note. Tabled errors = sum ≤18 for addition or products of single digit numbers for multiplication. Congruent intrusions = left operand appeared in decade of error or right operand appeared in unit of error. Incongruent intrusions = left operand appeared in unit of error or right operand appeared in decade of error. Non-intrusions = remaining tabled errors. Operation = sum instead of product or vice versa. Miscellaneous = any other error or no response.

Discussion

We found a smaller PSE in multiplication for auditory than Arabic format (cf. LeFevre, et al., 2001, Table 1, p. 280), but found the opposite for addition whereby the PSE was larger with auditory than Arabic stimuli. This triple interaction is difficult to reconcile with an encoding-based explanation of the Format × Problem Size interactions because we would expect encoding requirements to be similar for addition and multiplication. In this case, an effect at encoding would produce parallel effects for both operations, rather than the opposing Format × Problem Size interactions observed. The triple interaction therefore implies that format affected postencoding processes. Furthermore, the decomposition of mean RT into its ex-Gaussian components, μ and τ, demonstrated that the triple interaction arose entirely in connection with τ. Previous research has demonstrated that use of procedural strategies loads more on τ than on μ (Campbell & Penner-Wilger, 2006; Penner-Wilger, Leth-Steensen & LeFevre, 2002). Therefore the complex effects of format on RT observed here could reflect format-specific effects on strategy choice.

Such a proposal is consistent with the major assumptions of Dehaene’s triple-code model (Dehaene, 1992; Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2005; but see Pesenti, Thioux, Seron, & DeVolder, 2000; Thioux, Pillon, Samson, de Partz, Noël, & Seron, 1998; Venkatraman, Ansari, & Chee, 2005). The triple code model identifies two main processing pathways for arithmetic. One path involves rote verbal memory for arithmetic facts. The other path engages semantic elaboration processes to implement procedural strategies based on counting, decomposition, and so forth. Access to the semantic pathway may be especially efficient for Arabic input (Brysbaert, 2005). In this case, Arabic format could promote and facilitate application of semantic elaboration strategies, whereas auditory format could promote and facilitate direct retrieval from verbal memory.

These assumptions can neatly explain a salient and novel feature of our data. In the analysis of μ, there was a robust main effect of format with μ faster for auditory than Arabic. Also, τ presented a robust main effect of format but in the opposite direction with mean τ faster for Arabic stimuli than for auditory stimuli. If we assume that μ is influenced primarily by retrieval trials and that τ is influenced primarily by procedure trials (Campbell & Penner-Wilger, 2006), the opposing pattern of format effects could arise in the following way: With respect to retrieval trials (which load primarily on μ), the auditory format would facilitate direct retrieval from verbal memory, whereas the semantic pathway promoted by the Arabic format would interfere with direct verbal retrieval. With respect to procedure trials (which load primarily on τ), the Arabic format would facilitate semantic elaboration whereas the verbal retrieval promoted by the auditory format would delay or interfere with procedural strategies.

We can also suggest a tentative explanation for the three-way interaction observed for the overall RT analysis (i.e., the different Format × Size interactions for addition and multiplication). In the addition task, the PSE in mean RT was larger for auditory format than for Arabic format. For most educated adults, small addition facts (e.g., 2 + 3 = 5, 4 + 5 = 8) have high memory strength whereas larger additions (e.g., 7 + 9 = 16, 8 + 15 = 13) are less likely to be represented in memory and often must be solved via semantic elaboration (Campbell & Xue 2001; LeFevre, Sadesky, & Bisanz, 1996; Zbrodoff & Logan, 2005). If, relative to the Arabic format, auditory format promotes verbal retrieval this would facilitate direct memory retrieval for small additions or interfere with engaging the semantic processing that is often required for large addition problems.

For multiplication, the PSE in mean RT was larger for Arabic than auditory format. Relative to addition, educated adults rely more on direct memory retrieval for multiplication (Campbell & Xue, 2001). If Arabic format promoted semantic elaboration, relative to the auditory format, this would be most costly for large multiplication problems because large multiplication problems are particularly difficult to solve using procedural strategies (e.g., decomposition and repeated addition; cf. Campbell & Xue, 2001, p. 311). Consequently, even a small increase in the use of elaboration strategies would explain the increase of the PSE for multiplication in Arabic compared to auditory format.

In contrast to LeFevre et al. (2001), we did not find that the rate of operand intrusion errors for multiplication was greater with auditory than Arabic format, although there was a weak trend in that direction (162 for auditory vs. 146 for Arabic). The similar rates of intrusion errors with auditory and Arabic formats stands in contrast to repeated demonstrations that intrusions are much more likely with written verbal format (e.g., four × eight) than Arabic format. There have been two main theoretical accounts. Campbell (1994, 1997) argued that more intrusions with written word stimuli relative to Arabic stimuli occur because written number words activate reading processes that interfere with arithmetic fact retrieval. In contrast, Noël et al. (1997) argued that the effect has nothing to do with reading processes per se, but rather that written words activate phonological codes that interfere during response production. We would expect the auditory format to be at least as efficacious as the written word format with respect to activation of phonological codes. Consequently, if intrusions reflect phonological priming, and are not tied to reading processes per se, then we would expect a much higher rate of intrusions with auditory compared to Arabic stimuli. Thus, the similar rates of intrusions observed here with the auditory and Arabic stimuli are consistent with Campbell's (1994, 1997) conclusion that the higher rate of operand intrusions with written words is caused by numeral reading processes.

Conclusions

The current results add to the growing body of evidence that numerical surface form affects calculation processes per se and not only problem encoding processes (Campbell & Epp, 2005). We have suggested that differential effects of auditory numbers vs. Arabic numerals on performance of simple multiplication and addition could reflect differences in the probability that different solution pathways are activated. This proposal is consistent with the triple-code model proposed by Dehaene et al. (2005), which distinguishes between a verbal system that supports direct retrieval and an alternate pathway that supports semantic elaboration. These two systems are associated with distinct anatomical neural substrates (Dehaene et al., 2005). Future imaging studies could provide converging evidence that auditory and Arabic numerals promote arithmetic problem solving via the auditory and semantic pathways, respectively.

CHAPTER 4

Experimental research in cognitive arithmetic frequently relies on participants' self-reports to discriminate solutions based on direct memory retrieval from use of procedural strategies. Almost as long as this procedure has been used, there has been criticism that if retrieval processes are in fact automatic, participant self-reports would not be reliable. The concern has also been raised that self-reports are sensitive to instructional biases between experiments. Given concerns about the validity and reliability of strategy reports, Thevenot, Fanget, and Fayol (2007) developed the operand recognition paradigm as an objective measure of arithmetic strategies. Participants performed addition or number comparison on two sequentially presented operands followed by a speeded operand recognition task. Recognition times increased with problem size following addition but not comparison. Thevenot et al. argued that the complexity of addition strategies increases with problem size. A corresponding increase in operand recognition time occurs because, as problem size increases, working memory contains more numerical distracters. However, because addition is substantially more difficult than comparison, and difficulty increases with problem size for addition but not comparison, their findings could be due to difficulty-related task switching costs.

We repeated Thevenot et al. (Experiment 1) but added a control condition wherein participants performed a parity (odd or even) task instead of operand recognition. We replicated their findings for operand recognition but found robust, albeit smaller, effects of addition problem size on parity judgements. The results indicate that effects of strategy complexity in the operand recognition paradigm are confounded with task switching effects, which complicates its application as a precise measure of strategy complexity in arithmetic.

This chapter has been previously published. Reproduced here with kind permission from Springer Science+Business Media:

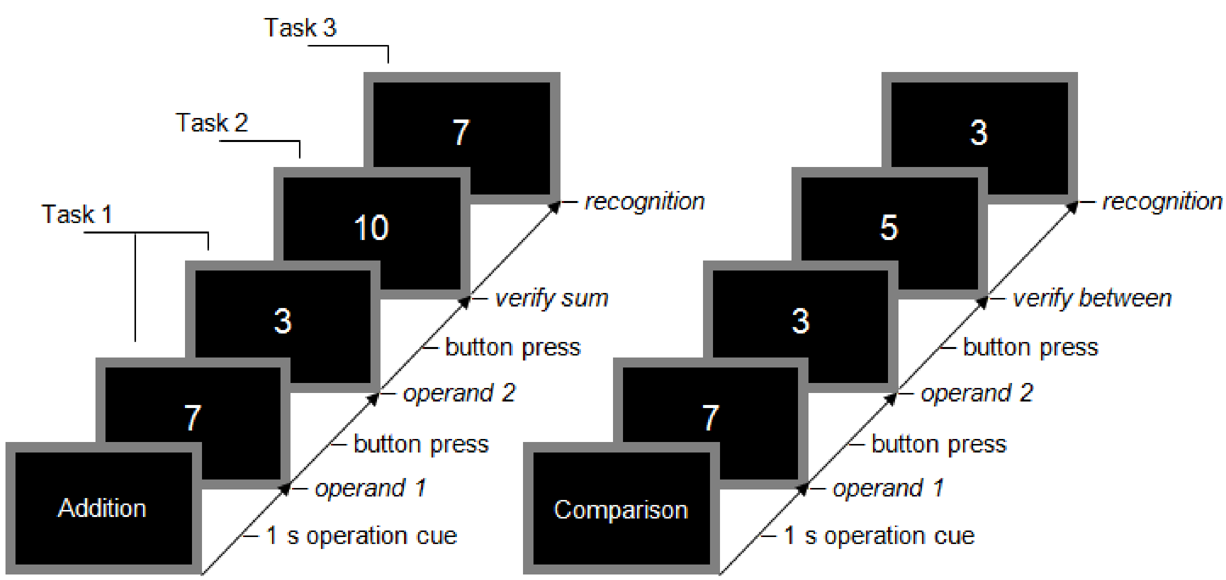
Metcalfe, A. W. S., & Campbell, J. I. D. (2010). Switch costs and the operand recognition paradigm. *Psychological Research, 74*, 491–498.

Switch Costs and the Operand Recognition Paradigm

The problem-size effect is the virtually ubiquitous phenomenon that the difficulty of simple arithmetic problems (e.g., 2 + 3, 6 × 9) increases as problem size increases. The problem-size effect has been recognized and studied systematically for over 80 years (e.g., Clapp, 1924), and continues to be a major focus of research on both children's and adults' elementary arithmetic skill (see Zbrodoff & Logan, 2005, for a review). For many years, the problem-size effect in adult performance was assumed to reflect slower and more error prone retrieval processes for large problems, presumably because large problems are encountered less often (Ashcraft, 1995). More recently, researchers have concluded that the problem-size effect in adults substantially reflects greater use of procedural strategies (e.g., counting, decomposition) for large problems (e.g., Campbell & Xue, 2001; LeFevre, Sadesky, & Bisanz, 1996; Imbo, Vandierendonck & Rosseel, 2007). Measurement of strategies in these studies has relied predominantly on verbal self-reports. The assumption is that procedural strategies (e.g., 4 + 9 = 4 + 10 - 1= 14 - 1 = 13) generate a series of verbal products in working memory that are available to introspection. These conscious products provide a basis for participants to identify and retrospectively report their solution strategy. Nonetheless, the subjectivity of strategy self-reports raises concerns about their validity and reliability. Indeed, Kirk, and Ashcraft (2001) demonstrated that adults' strategy reports for simple arithmetic are very sensitive to instructional biases. Consequently, finding objective measures that provide converging evidence for procedural strategies is highly desirable.

To this end, Thevenot and colleagues (Thevenot, Barrouillet, & Fayol, 2001; Thevenot, Fanget, & Fayol, 2007) introduced the operand-recognition paradigm as an objective index of procedural strategies. The paradigm was developed in the context of adults' arithmetic performance but has the potential for a wide variety of applications in behavioural and imaging research (Thevenot et al., 2007). The operand-recognition paradigm employed a series of three tasks on each trial (see Figure 4-1). Each trial began with a cue for the operation to be performed on the current trial (add or compare). After the cue, participants sequentially viewed a pair of numerical operands (henceforth Task 1). The operand pairs varied in numerical size: small (single digits with a sum ≤ 10), medium (single digits with a sum > 10), and large (double-digit operands). Presentation duration of each of the two operands was controlled by the participant via a button press. Task 1 was followed immediately by Task 2. In Task 2, an answer was presented and participants verified if the proposed answer was true or false. On trials for which the operation was cued as "addition", participants verified the sum of the two operands. On trials for which the operation was cued as "comparison", participants verified if the proposed answer numerically fell between the two operands. Following the verification response to Task 2, a fourth number was presented for a recognition decision on one third of trials (henceforth Task 3). Participants completed the recognition task by verifying if the proposed number matched either of the problem operands presented in Task 1. For Task 2 and Task 3, the proposed answer was correct for half of the trials.

Figure 4-1. Trial Event Sequence in the Operand-recognition Paradigm (Thevenot et al., 2007)



Thevenot et al. (2001, 2007) argued that the speed and accuracy of the recognition decision would be diagnostic of the complexity of the cognitive operations performed during the addition and comparison operations. For addition, complexity was expected to increase with problem size. That is, small problems with a sum ≤ 10 are usually solved by direct retrieval, whereas medium, and especially large, problems would usually involve procedural strategies that generate additional elements in working memory. For example, for the problem 2 + 3, direct retrieval of the answer 5 would entail a single step, whereas a procedural strategy for 28 + 13 (e.g., 28 + 10 = 38 + 3 = 41) would generate several intermediate results. Each additional element in working memory, along with decomposition of the operands, potentially interferes with a subsequent recognition decision about the original operands. Therefore, recognition performance was expected to decrease as addition problem size increased. In contrast, when Task 2 required comparison (i.e., deciding if the proposed answer fell between the two Task 1 operands), there would not be much working memory clutter generated and problems operands would be processed similarly regardless of problem size. Consequently, recognition following comparison should be relatively unaffected by problem size.

Indeed, Thevenot et al. (2007, Experiment 1) found that mean RT for operand-recognition was 1180 ms, 1358 ms, and 1562 ms following small, medium, and large addition problems, respectively, whereas comparison size had little effect on subsequent recognition performance (small 1127 ms, medium 1157 ms, large 1185 ms). This pattern of results fit well with the hypothesis that recognition performance would vary with strategy complexity. Thus, Thevenot et al. concluded that operand-recognition time provided an objective index of the complexity of the addition solution strategy.

The Operand-recognition Paradigm and Task Switch Costs

Recently, however, research unrelated to the operand-recognition paradigm demonstrated phenomena that potentially complicate its interpretation. Campbell and Metcalfe (2008) investigated effects of context (e.g., addition vs. comparison) on speed to name single Arabic digits. Consistent with previous neuropsychological research (Cohen & Dehaene, 1995), digit naming was more efficient in the context of comparison blocks than number-fact retrieval blocks. Incidental to this finding, however, Campbell and Metcalfe also observed RT costs on digit naming associated with the difficulty of the preceding problem. Specifically, naming RT was slower following a large, relatively difficult arithmetic problem relative to a small problem, but the size of comparison problems had no effect on subsequent naming RT. This suggests that the task switch from a large arithmetic problem to digit naming incurred an RT cost. Much research has demonstrated RT and accuracy costs to switch between tasks (Monsell, 2003). Furthermore, it is often found to take longer to switch from a difficult task than from a relatively easier task (Allport, et al., 1994; Arbuthnott, 2008; Yeung & Monsell, 2003). A difficult task set is more likely to remain highly activated and interfere with initiating a new task (Monsell, et al., 2000).

Differential switch costs following addition versus comparison raise a possible concern with the operand-recognition paradigm: Effects of Task 2 (i.e., addition or comparison) on recognition RT could reflect switch costs rather than interference with recognition processes per se. The within-trial task transitions in the operand-recognition paradigm are structurally similar to the conditions under which switch costs are observed when tasks switch in a continuous sequence of trials (Monsell, 2003). Furthermore, the pattern of task-switch costs on digit naming observed by Campbell and Metcalfe (i.e., an Operation × Size interaction whereby RT increased with problem size following arithmetic but not comparison) mimics the results for recognition RT observed by Thevenot et al. (2007, Experiment 1). The magnitude of the Operation × Size interaction observed in digit naming by Campbell and Metcalfe was as large as 42 ms (Experiment 3), but was much smaller than the average 191 ms effect on recognition RT observed by Thevenot et al. (2007, Experiment 1). Nonetheless, switch costs might be larger in the operand-recognition paradigm. In the latter, the recognition cue followed the arithmetic response immediately and unpredictably, whereas task switching was predictable and measured between trials in the Campbell and Metcalfe experiments. These differences could result in larger switch costs in the operand-recognition paradigm (Monsell, 2003).

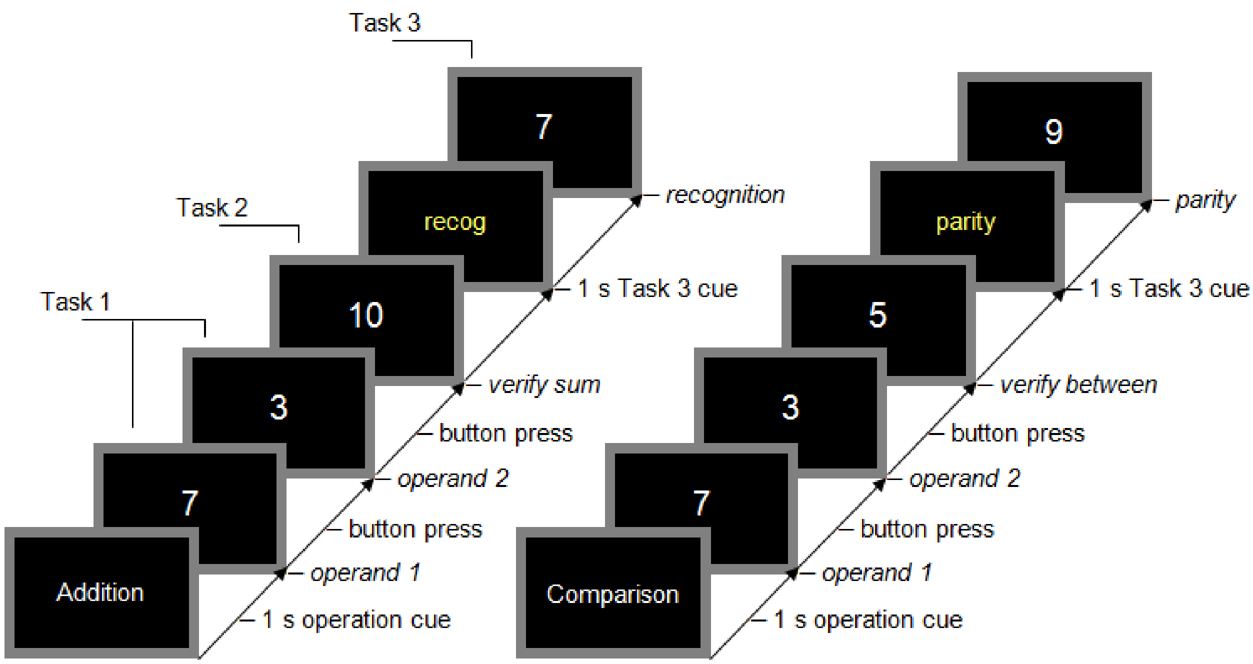
To investigate the possibility of switch costs in the Thevenot et al. paradigm, we based a new experiment on Thevenot et al. (2007, Experiment 1) but designed a novel Task 3 condition in which a parity judgement (e.g., 5 = odd/even?) was performed on the same numerical stimuli used for the recognition task (see Figure 4-2). The parity task should be sensitive to task switch costs and thereby served as a task switching control condition.[10] If switch costs were responsible

---

[10] An earlier experiment in our lab found no switch costs using a letter-comparison control task in which participants decided if a presented letter preceded or followed the letter E. This was a poor task-switch control, however, because the numerical task set (i.e., addition or number comparison) would not compete with letter comparison. In contrast, like operand recognition, parity judgments following addition or comparison required participants to switch number-processing task sets.

for the findings of Thevenot et al. we would expect to observe a similar Operation × Problem Size interaction for both parity and operand-recognition. In contrast, if arithmetic strategies are responsible for recognition costs as Thevenot et al. proposed, we expected Operation (addition vs. comparison) × Problem Size (small, medium, and large) effects for recognition but not parity.

Figure 4-2. Trial Event Sequences in the Current Experiment



Note. The figure depicts addition followed by recognition at Task 3 and comparison followed by parity at Task 3, but all four Operation × Task 3 combinations were tested.

Method

Participants

Twenty-four volunteers (15 female; 22 right-handed; mean age 19 years) were recruited from the University of Saskatchewan Introductory Psychology participant pool. Each participant received course credit for participation.

Apparatus, Stimuli, and Design

The experiment was performed using E-prime software on a MS Windows-based computer. Participants were seated approximately 60 cm from the monitor. RTs accurate to ±1 ms were collected via button box. Stimuli appeared at a central fixation point, were approximately 4 mm wide × 6 mm high, and were based on the 24 number pairs (8 small, 8 medium, and 8 large) used by Thevenot et al. (2007). Each trial was comprised of two or three tasks: Task 1 involved sequential presentation of an operand pair for the purpose of addition or comparison. “Small” pairs included 5 2, 5 3, 6 2, 6 3, 6 4, 7 2, 7 3, 8 2, “medium” pairs included 7 5, 8 4, 8 5, 8 6, 9 4, 9 5, 9 6, 9 7, and “large” pairs were 28 13, 35 16, 36 17, 38 16, 39 26, 43 18, 43 19, 47 16. The first number in each pair was always the numerically larger of the two. Task 2 consisted of the presentation of proposed answers to either addition or comparison solutions. For each operation, each Task 1 pair was presented twice, once followed by a true answer and once by a false answer. For addition, true answers were the sum of the two addends in the pair and false answers were randomly selected to correspond to the sum ±1. On comparison trials, true answers were randomly selected from between the two problem operands (for large problems the answer always shared the decade with one operand) whereas false answers were randomly selected to be one number above the larger or one number below the smaller operand[11].

Task 3 included three conditions: Recognition, parity, and no Task 3. The recognition condition was the same as Thevenot et al. (2007, Experiment 1) and required participants to decide if a proposed digit matched one of the two problem operands presented in Task 1. On true recognition trials the stimulus matched the first Task 1 operand half the time and the second Task 1 operand on the other half. On false recognition trials the stimulus corresponded to the first Task 1 operand ±1 half the time and the second operand ±1 on the other half. When Task 3 was

[11] For the problem pair 7 5, Task 2 trials were always paired with a false comparison due to a coding error; this affected .01% of Task 2 trials and had no practical consequences for the recognition or parity tasks. Also, because false Task 2 addition answers were ±1 from the correct sum participants could use addition parity rules to solve Task 2 addition problems (Lemaire & Reder, 1999). Future research might wish to avoid this possibility, but the current experiment was designed to be a close replication of Thevenot et al. (2007, Experiment 1) and adopted their rules for false answer stimuli.

a parity judgment, participants decided if a presented number was odd (1, 3, 5, etc.) or even (2, 4, 6, etc.). Small, medium, and large parity task stimuli were selected randomly with replacement from the set of numbers used for the recognition task. For both versions of Task 3, stimuli were selected with the constraint that they could not match the proposed arithmetic solution presented in Task 2.

The 3 (size) × 2 (operation) × 2 (truth) design multiplied by the eight operand pairs in each size condition yielded 96 different problem-answer combinations of Task 2 stimuli. All 96 combinations were tested once followed by a Task 3 recognition trial, once followed by a Task 3 parity trial, and once without Task 3. Within each of the 12 Size × Operation × Truth cells, half of the Task 3 trials were true and the other half were false. The pattern of true and false Task 3 stimuli was yoked for the comparison and addition operations so that participants made identical Task 3 decisions in the context of both addition and comparison problems. Within these constraints, operand pairs were assigned randomly to conditions and the same assignment was used for all participants. The resulting 288 trials were presented in a single block with trial order independently randomized for each participant.

Procedure

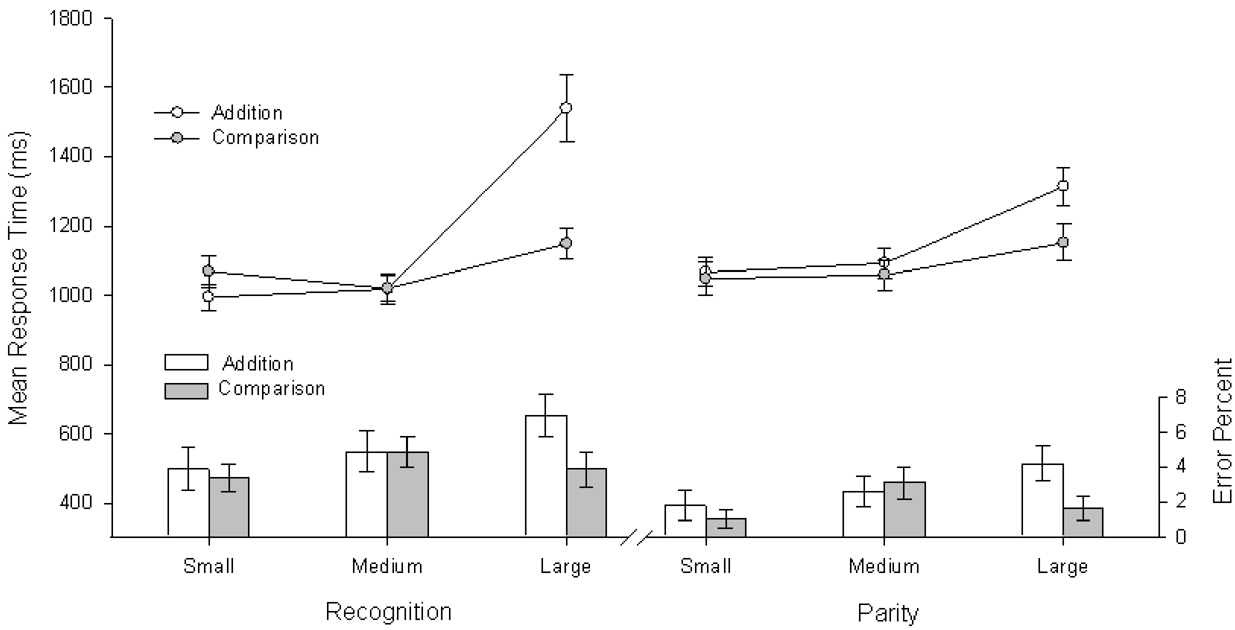
Participants were tested individually in a single session lasting about 50 minutes. Instructions indicated that participants would be presented with 144 comparison problems and 144 addition problems in a random order and that a recognition or parity decision would follow on two-thirds of trials. For practice, participants received three operand recognition and three parity trials based on stimuli not used for experimental trials. Each trial (see Figure 4-2) began with a 50 ms blank followed by 1 s presentation of the “COMPARISON” or “ADDITION” cue. The operation cue was immediately followed by the first problem operand. Participants were instructed to press either button used for the verification task (one labeled “YES” the other labeled “NO”) when they were ready to view the second problem operand. Participants were instructed to again press either button when they were ready to view the proposed answer. Task 2 instructions advised participants to press the “YES” key if the proposed answer was true or the “NO” key if it was false. On two-thirds of trials, immediately following the response to Task 2 the respective cue “Recog” or “Parity” was presented for 1 s in yellow font. The cue was immediately followed by presentation of the respective recognition or parity answer for verification. On recognition trials participants were instructed to press the YES key if the proposed number matched either of the problem operands or NO if it did not. On parity trials, instructions and assignment of “odd” and “even” to YES and NO was counterbalanced. Timing began when each stimulus appeared (synced with the vertical retrace) and was terminated by the button press response. Assignment of left and right keys to YES and NO responses was counterbalanced across participants. Feedback about performance was provided on the six practice trials only. Participants had the opportunity for a short break at intervals of 72 trials.

Results

Task 3 (recognition or parity)

Mean correct Task 3 RTs were analyzed using a participant-based 2 × 2 × 3 repeated-measures ANOVA including the factors Task 3 (recognition or parity), Task 2 (add or compare), and problem size (small, medium, large). The error rate for Task 3 performance was 3.5%. RTs >3 standard deviations from each grand cell mean (1.9% of trials) and RTs following incorrect Task 2 responses (2.5%) were not included in the analysis. Figure 4-3 depicts the three-way interaction [$F(2, 46) = 6.83$, $MSE = 25756$, $p = .003$]. Task 3 recognition RTs substantively replicated Thevenot et al. (2007), inasmuch as mean RTs were much longer following large additions relative to large comparisons [+390 ms, $t(23) = 5.07$, $SE = 76.9$, $p < .001$]. Most important for our current experimental hypothesis, Task 3 parity RTs also were significantly longer following large additions relative to large comparisons [+162 ms, $t(23) = 2.43$, $SE = 66.6$, $p = .02$]. The magnitude of the effect on parity was smaller than the effect on recognition, however, giving rise to the three-way interaction. An item-based analysis confirmed longer T3 RTs following large additions than large comparisons both for recognition [+355 ms, $t(7) = 3.71$, $SE = 95.7$, $p = .008$] and parity [+201 ms, $t(7) = 2.48$, $SE = 81.0$, $p = .04$]. As Figure 4-3 shows, Task 3 RT was not differentially affected by addition vs. comparison following small or medium problems. Thus, both participant-based and item-based analyses confirmed that parity RT increased following large addition problems relative to large comparison problems. This effect implies a significant contribution of difficulty-related switch-costs to performance of Task 3.

Figure 4-3. Response Time (in milliseconds) and Error Percentage on Task 3 Recognition and Parity Decisions as a Function of Operation (addition and comparison) and Problem Size (small, medium, and large)



Note. Error bars are ±1 *SE.*

With respect to Task 3 errors, a participant-based Task 3 × Task 2 × Size repeated-measures ANOVA indicated more errors for recognition (4.7%) than for parity (2.4%) [$F(1, 23) = 10.95$, $MSE = 34.81$, $p = .003$], more errors after addition (4.1%) than comparison (3.0%) [$F(1, 23) = 5.12$, $MSE = 16.55$, $p = .033$], and that errors increased as a function of Problem Size (small 2.5%, medium 3.9%, large 4.2%) [$F(2, 46) = 4.23$, $MSE = 17.33$, $p = .021$]. The main effects of Task 2 operation and Size were qualified by a Task 2 × Size interaction [$F(2, 46) = 4.66$, $MSE = 13.40$, $p = .014$] that occurred because Task 3 errors increased with Size following addition (small = 2.9%, medium = 3.8%, large = 5.6%) more so than following comparison (2.2%, 4.0%, 2.7%). As Figure 4-3 shows, the Task 2 × Size effect reflected more errors following large addition than large comparison for both recognition [+3.1%, $t(23) = 2.30$, $SE = 1.36$, $p = .03$] and parity [+2.6%, $t(23) = 2.46$, $SE = 1.06$, $p = .02$]. Item-based analyses confirmed more parity errors following large addition than large comparisons [+2.7%, $t(7) = 2.0$, $SE = 1.29$, $p = .08$], but the corresponding test for recognition was weaker [+3.1%, $t(7) = 1.7$, SE

= 1.98, p = .14]. There was no evidence for a three-way interaction ($F < 1$). Thus, both the participant-based and item-based errors analyses confirmed increased parity errors following large addition problems relative to large comparison problems. The error and RT analyses thereby provided converging evidence that difficulty-related switch-costs affected Task 3 performance.

Task 1 (operand processing) and Task 2 (addition or comparison)

Table 4-1 includes mean RTs for operand processing (Task 1) and mean correct RTs and error percentages for the verification decision (Task 2). Following Thevenot et al. (2007), total solution time (i.e., time to view first operand + second operand + verification) was analyzed using a Task 2 (addition, comparison) × Size (small, medium, large) ANOVA.[12] Incorrectly answered Task 2 trials (2.5%) were omitted from the analysis. RTs > 3 standard deviations from grand cell means were discarded as outliers (1.5% in Task 2, 1.7% in Task 1). Addition (3756 ms) was slower on average than comparison (3389 ms) [$F(1, 23) = 15.00$, $MSE = 324479$, p = .001] and RT increased with size (2886 ms, 3168 ms, 4663 ms) [$F(2, 46) = 95.31$, MSE = 459234, $p < .001$]. As expected (cf. Thevenot et al., 2007), size interacted with operation [$F(2, 46) = 62.04$, $MSE = 302832$, $p < .001$] such that addition RTs (small 2634 ms, medium 3071 ms, large 5564 ms) increased more with problem size than did comparison RTs (3139 ms, 3265 ms, 3762 ms).

[12] We refer to the component stages of addition and comparison as Task 1 and Task 2, but because the Task 2 decision stage necessarily refers to the processes and representations generated during Task 1, it is reasonable to assume that they constitute an integrated task set and that task difficulty is properly indexed by total solution time (i.e., Task 1 + Task 2 RT).

Table 4-1. *Mean Response Times (in milliseconds) for Task 1 and Task 2 and Mean Error Percentages for Task 2*

	Task 1		*Task 2*	
	RT		RT	Errors
Problem Size	First Operand	Second Operand		
Addition				
Small	965	822	846	1.3
Medium	961	1091	1019	2.0
Large	1226	2964	1374	6.3
Comparison				
Small	1109	873	1157	1.6
Medium	1115	966	1184	2.0
Large	1214	1163	1385	1.9

A Task 2 operation × Problem Size ANOVA performed on Task 2 error percentages indicated a main effect of Task 2 operation (addition 3.2%, comparison 1.8%) [$F(1, 23) = 4.82$, $MSE = 13.81$, $p = .038$], a main effect of Size (small 1.4%, medium 2.0%, large 4.1%) [$F(2, 46) = 12.85$, $MSE = 7.26$, $p < .001$], and a Task 2 × Size interaction [$F(2, 46) = 42.44$, $MSE = 7.70$, $p < .001$]. The interaction reflected a greater increase in errors with problem size for addition (small 1.3%, medium 2.0%, large 6.3%) than for comparison (1.6%, 2.0%, 1.9%). The analyses of Task 1 and Task 2 RTs and Task 2 errors substantively replicated Thevenot et al. (2007), demonstrating that problem difficulty increased more with problem size for addition than comparison, especially for the large addition problems. As these Task 2 performance differences were associated with corresponding Task 2 × Size effects on both recognition and parity at Task 3, the results strongly suggest that Task 3 was subject to difficulty-related task-switch costs.

Discussion

As Thevenot et al. (2007) found, operand recognition latency substantially increased following large addition problems but increased only slightly following large comparison problems. Here we demonstrated that parity judgments too were more difficult following large additions relative to large comparisons. This effect emerged both in RT and error measures using both participant- and item-based analyses. Thus, the statistical evidence was clear that parity performance was disrupted following large addition problems relative to large comparison problems.

These results have important implications for the operand recognition paradigm. Thevenot et al. (2001, 2007) argued that operand recognition time increased with addition problem size, but not comparison problem size, because computational complexity increases with size for addition but not comparison. As computational complexity increases, working memory becomes increasingly cluttered with numerical distracters, and operand representations are degraded by the operations performed on them, thus interfering with subsequent operand recognition. Unlike operand recognition judgments, however, the parity judgments could be made independently of the specific stimuli presented and without reference to the working memory contents generated during operand processing (Task 1) or verification (Task 2). Nonetheless, the pattern of RTs and errors on the parity task mirrored the difficulty of the corresponding operation (i.e., addition compared to comparison). This strongly suggests that the effects on parity judgments reflected difficulty-related task-switching costs. As both parity and operand recognition entailed a number-processing task switch following Task 2, we would similarly expect operand recognition to encounter difficulty-related switch costs. The results therefore imply that the effects of operation and problem size on operand recognition included switch costs rather than only effects owing to the complexity of the cognitive operations performed during calculation as assumed by Thevenot et al. (2007).

It is important to emphasize, however, that operand recognition presented much larger RT costs following large additions (+390 ms relative to large comparisons in the participant-based analysis) than did parity (+162 ms). Thus, based on the current results (and assuming equivalent switch costs for parity and recognition), approximately 60% of the effect would not be attributable to switch costs. Furthermore, specific features of the recognition costs are difficult to explain as switch costs. Thevenot et al. (2007) found that recognition costs following addition relative to comparison were larger for the second operand than for the first (442 ms and 312 ms, respectively). Here, similarly, recognition RT costs following large addition relative to large comparison were larger for the second operand (+631 ms) than for the first (+140 ms)[13]. Thevenot et al. proposed that performing addition (e.g., 28 + 13) more often involves decomposing the second operand than the first one (e.g., 28 + 10 + 3); therefore, memory traces of the second operand are weaker than the first. This feature of performance in the operand recognition paradigm is difficult to account for in terms of switch costs and supports the conclusions of Thevenot et al. (2007) that operand recognition times are sensitive to the specific arithmetic strategies performed prior to the recognition decision.

Thevenot et al. (2007) observed slower recognition RT following addition than comparison for both large and medium problems, whereas the operation effect appeared only for large problems here. In their study, however, total addition RT increased by +745 ms from small to medium problems and by +1383 ms from medium to large. Here, the RT increase from small to medium additions was only +437 ms, compared to an increase of +2493 ms from medium to large. Thus, our participants found the small and medium additions to be more similar in difficulty, especially compared to large problems. Consequently, in the present study we would expect similar switch costs in connection with small and medium problems, especially compared to large additions.

Finally, switch costs in the present study (as estimated by longer parity RT following large additions relative to large comparisons) were much larger (+162 ms) than the switch costs of up to +42 ms observed by Campbell and Metcalfe (2008). It is not surprising, however, that task-switch effects would be larger in the current study. First, the switch costs in Campbell and Metcalfe were measured on single-digit naming RTs. Establishing the task set for simple digit naming is presumably very straightforward, potentially mitigating switch costs. Second, from a switch-cost standpoint, the multi-digit addition trials used in our current operand recognition paradigm experiment were much more difficult than the single-digit arithmetic problems used in our previous research. Consequently, the relatively larger effects here could be owed to difficulty-related switch costs. Third, in Campbell and Metcalfe (2008), the specific task switch was 100% predictable and reoccurred consistently 27 times per block of trials. In contrast, in the

[13] As the parity task did not afford analysis in terms of first versus second operand we did not include this as a factor in our main analysis.

operand recognition experiment here, the switch required at Task 3 (recognition, parity, or no Task 3) was variable. Given all this, it would be reasonable to expect larger switch costs in the current study compared to Campbell and Metcalfe (2008).

Conclusions

Our results indicate that effects of strategy complexity in the operand recognition paradigm are confounded with task switching effects. Consequently, the operand recognition paradigm in its current form is not an unbiased measure of strategy complexity. To address this problem, switch costs on recognition decisions might be controlled by making two successive responses as part of Task 3 performance (e.g., two successive recognitions). The first Task 3 judgment could absorb the task switch (Campbell & Metcalfe, 2008; Monsell, Sumner & Waters, 2003), leaving the second judgment free of switch costs. Future development of the operand recognition paradigm could also seek to manipulate the recognition target to match specific numerical elements expected to be generated by a given strategy (e.g., 10 as a recognition target following solution of 28 + 13 by decomposition into 28 + 10 + 3). The use of that strategy would be signalled by increased false positives or increased latencies relative to control problems in the operand recognition task. The operand recognition paradigm may yet provide a valuable objective diagnostic tool for assessing arithmetic strategies despite the evidence that, in its current form, interpretation of recognition performance is complicated by task-switching costs.

CHAPTER 5

A critical issue for cognitive arithmetic research is valid and precise measurement of strategy use (i.e., direct memory retrieval vs. procedural strategies). Verbal self-reports have been the most common measure, but Thevenot et al. (2007) introduced an ostensibly more objective method to measure arithmetic strategy. In their operand recognition paradigm (ORP), speed of recognition memory for problem operands immediately after solving a problem provides an index of strategy use. However, given the results described in Chapter 4 that cognitive set task switching might be responsible for as much as half of the observed effects previously attributed to strategy use, it was unclear how to interpret the results of the operand recognition method. On one hand, the ORP appeared to be a solution to the concerns over of validity and reliability of self-report measures of basic arithmetic strategy use, but on the other hand, the results from the parity control task (Chapter 4) strongly suggested that the paradigm was confounded with arithmetic task difficulty. The work in the current chapter was undertaken to address this uncertainty in how to interpret operand recognition data.

Previous research has indicated greater use of procedural strategies for simple addition than multiplication, but in two experiments ORP recognition performance was constant for simple addition and multiplication. This was true both when operation was manipulated between-participants in Experiment 1 and within-participants in Experiment 2. This null result appears to represent a basic failure of the paradigm to detect a robust and widely recognized effect of operation on procedure use. The findings indicate, once and for all, that the ORP is not a reliable substitute for verbal reports to measure strategy choices.

This chapter is currently under review in an academic journal:

Metcalfe, A. W. S., & Campbell, J. I. D. (2010). *Adults' strategies for simple addition and multiplication: Verbal self-reports and the operand recognition paradigm.* Manuscript submitted for publication.

Verbal Self-Reports and the Operand Recognition Paradigm

Strategy use for adults' basic arithmetic has received much experimental investigation in recent years (Campbell & Fugelsang, 2003; Campbell & Albert, 2009; Campbell & Xue, 2001; Dehaene, Piazza, Pinel, & Cohen, 2003; Geary & Wiley, 1991; Kirk & Ashcraft, 2001; LeFevre et al., 1996b; LeFevre, Sadesky, & Bisanz, 1996; Metcalfe & Campbell, 2007; Metcalfe & Campbell, 2010; Thevenot, Castel, Fanget, & Fayol, in press; Thevenot, Fanget, & Fayol 2007). The most common method for assessing strategy use has been participants' verbal self-reports. The validity of self-reported strategies has been challenged, however (see Kirk & Ashcraft, 2001), and proponents of the method have subsequently attempted to demonstrate some degree of validity (Campbell & Austin, 2002; Campbell & Penner-Wilger, 2006; Smith-Chant & LeFevre, 2003). More recently, Thevenot and colleagues (Thevenot, Barrouillet, & Fayol, 2001; Thevenot et al., in press; Thevenot et al., 2007) have proposed a new method for objectively measuring strategy use, the *operand recognition paradigm* (ORP). The primary goal of the present research was to compare ORP results with verbal self-reports for simple addition and multiplication to determine the extent to which they converge or diverge in their implications about strategy choice.

The Operand Recognition Paradigm

Performance of simple arithmetic (e.g., 9 + 6, 7 × 8) is generally based on one or the other of two alternative types of strategies: Direct retrieval from long-term memory or a procedural strategy based on transformation or counting (e.g., Campbell & Xue, 2001). The basic premise of the ORP is that people's ability to recognize problem operands (9 + 6) after they have answered the problem will be affected by the strategy used to solve the problem. For directly retrieved answers, operand memory is expected to be fast and accurate because the retrieval strategy does not generate extraneous working memory contents nor involve dividing attention among intermediate steps of the solution process (see Thevenot et al., 2007, pp. 1345-1346). In contrast, for problems solved via procedural strategies (e.g., 9 + 6 = 10 + 6 - 1 = 15), division of attentional resources and increased working-memory interference is expected to produce relatively poor memory for problem operands.

Thevenot et al. (2001) originally examined the ORP on double-digit comparison (does a target numerically fall between two operands?), addition, and subtraction problems. Participants were presented with an operation cue (comparison, addition, or subtraction), followed by two problem operands presented sequentially, and then a proposed answer to verify as true or false. On one-third of trials, participants subsequently received a recognition probe that did or did not match one of the problem operands. The display sequence on a recognition trial would proceed thus: Addition/39/16/55/41 (Thevenot et al., 2001, p. 602). In this example, 55 is the correct addition answer to 39 + 16 and 41 is a no-match recognition probe (either 39 or 16 would be a matching probe). Thevenot et al. demonstrated that recognition response time (RT) after addition and subtraction was significantly longer than recognition RT following comparison problems. They interpreted this as evidence that elaborated solution strategies for double-digit arithmetic problems (e.g., –) interfered with operand memory relative to double-digit comparison problems that do not require complex solution strategies.

Thevenot et al. (2007) subsequently applied the ORP to the question of the prevalence of multiple solution strategies in adults' single-digit addition and double-digit addition. Based primarily on the work of LeFevre, Sadesky, and Bisanz (1996; but see Thevenot et al., 2007, pp. 1344 – 1345 for a more thorough review) they predicted that non-retrieval strategies would increase with problem size for addition but not for comparison. These differences in strategy use would result in a problem-size effect on recognition RTs following addition but not following comparison. Problems were divided into "small" (sums < 10), "medium" (single-digit problems with sums > 10), and "large" (double-digit addends). As predicted, the results of Experiment 1 indicated that recognition RT increased with size following addition problems but not following comparison problems. Furthermore, in a second experiment they demonstrated that recognition RT for medium addition problems only differed from that of small problems in a low-skilled arithmetic group. This second result further supported the interpretation that recognition RT was sensitive to strategy use because skill-level should be positively correlated with retrieval use for more difficult, but practiced, single-digit problems.

Campbell and Metcalfe (2008), however, identified complications for this interpretation of the ORP. They studied the effect of context (e.g., addition vs. comparison) on speed to name single Arabic digits. Naming RT was slower following a large addition or multiplication problem relative to a small problem, but the size of comparison problems had no effect on subsequent naming RT. Campbell and Metcalfe interpreted this as evidence that the task switch from an arithmetic problem to digit naming was responsible for the RT cost. In concordance with Thevenot et al.'s (2007) interpretation of comparison task difficulty (i.e., invariant over size), the effect of size was not observed following comparison problems (Campbell & Metcalfe, 2008, p. 235). This raised the possibility that effects of arithmetic problem size in the ORP reflected difficulty-related switch costs rather than strategy use.

Metcalfe and Campbell (2010) subsequently investigated the possibility that difficulty-related switch-costs were responsible for the Operation × Size interaction observed by Thevenot et al. (2007). Participants received the standard ORP (Thevenot et al., 2007; Experiment 1) with one critical change: In addition to the operand recognition probe following one-third of addition and comparison trials, participants also received a parity control task on another one-third of trials following addition or comparison. On the parity task participants were asked to classify the target number as odd or even. Key to this manipulation was that the parity task did not require reference to the problem operands. Similar to Thevenot et al., Metcalfe and Campbell found an Operation × Size interaction on recognition RT whereby recognition RT for large addition problems was significantly longer than recognition RT for large comparison problems (+390 ms). But more importantly they also observed a similar, albeit smaller, Operation × Size effect on parity RT (+162 ms).

Metcalfe and Campbell (2010) interpreted this result as evidence that operand recognition RT was directly affected by the difficulty of the preceding problem. Indeed, much research has demonstrated RT and accuracy costs to switch between tasks (Monsell, 2003). Furthermore, it is often found to take longer to switch from a difficult task than from a relatively easier task (Allport, Styles, & Hsieh, 1994; Arbuthnott, 2008; Yeung & Monsell, 2003). Thus, differential switch costs following addition versus comparison raise a possible concern with the ORP. Indeed, up to 40% of the ORP effect they observed was potentially attributable to switch costs.

While the results of Metcalfe and Campbell (2010) raised doubts about how precisely ORP measures strategy use, Thevenot et al. (in press) presented evidence that the ORP detected variation in strategy use not detected by verbal reports. Specifically, Thevenot et al. (Experiment

2) examined operand recognition on subtraction problems of three sizes (small: single digit problems; medium: minuends from 11-17 and single digit subtrahends; large: double-digit problems) contrasted with a comparison control task based on Thevenot et al. (2007). The experiment was comprised of three phases. In the first phase participants completed a test of arithmetic skill. In the second phase they completed the main operand recognition task described above. In Phase 3 participants solved the same subtraction problems from Phase 2 and self-reported the strategy used following each problem. For large and small problems, increased reports of procedural strategies for subtraction corresponded with longer recognition RT relative to comparison for both low and high-skilled participants. Thus, recognition RT and strategy reports converged with respect to strategy use for small and large subtractions both for low and high skill participants. However, for medium problems, both low and high skill participants reported similar rates of procedure use, but operand recognition times implied that low-skill participants used procedures more for medium problems than small problems, but high-skill participants did not. Thevenot et al. interpreted these result as evidence that operand recognition was more sensitive to procedure use than verbal self-reports.

The Present Experiment

The currently available evidence is mixed with respect to the relative validity and precision of the ORP vs. verbal self-reports to measure strategy use in arithmetic. The primary goal of the present research was to compare ORP results with verbal self-reports for simple addition and multiplication to determine the extent to which they converge or diverge in their implications about strategy choice. Additionally, these experiments were the first to examine ORP performance using the arithmetic production task in which participants verbally produce an answer, as opposed to the true-false verification task used in previous ORP research.

There are several sources of converging evidence that simple addition (e.g., 8 + 9) relies on procedural strategies more than does simple multiplication (e.g., 8 × 9). Verbal self-reports consistently demonstrate that large addition problems especially, (e.g., 7 + 9) often are solved by procedures that require semantic mediation (e.g., 10 + 7 - 1 = 16; see Campbell & Alberts, 2009; Campbell & Xue, 2001; Kirk & Ashcraft, 2001; LeFevre, et al., 1996a). In contrast, simple multiplication relies largely on direct verbal retrieval and rarely involves semantic elaboration

(Campbell & Epp, 2005; Campbell & Xue, 2001; Kirk & Ashcraft, 2001; LeFevre et al., 1996b). This view is reinforced by brain imaging research showing that addition problems are more likely than multiplication problems to result in left parietal activation implicated in semantic elaboration strategies (Dehaene, Piazza, Pinel, & Cohen, 2003; Pesenti, Thioux, Seron, & De Volder, 2000; Stanescu-Cosson, et al., 2000; Venkatraman, Ansari, & Chee, 2005). Furthermore, lesions in language-related areas, which are implicated in long-term memory storage of arithmetic fact, impair multiplication ability while leaving addition performance relatively spared (Dehaene & Cohen, 1997; Dehaene, Piazza, Pinel, & Cohen, 2003; van Harskamp & Cipolotti, 2001). Given the evidence that addition relies more on procedures than multiplication, then recognition performance in the ORP should be slower following addition than following multiplication if the ORP is sensitive to the use of procedural strategies.

In contrast, the hypothesis that the ORP is largely sensitive to difficulty-related switch costs leads to the prediction of little effect of operation on ORP performance. Previous research testing simple addition and multiplication indicates a small, overall RT and accuracy advantage for addition compared to multiplication (e.g., Campbell, 1994: -50 ms; Campbell & Arbuthnott, 2010: -54 ms; Campbell, Parker, & Doetzel, 2004: - 36 ms; Campbell & Xue, 2001: -81 ms; Metcalfe & Campbell, 2008: -45 ms; see also Geary, Widaman & Little, 1986; Hecht, 1999). Thus, even though addition relies more on procedural strategies, multiplication is slightly slower and error prone overall relative to simple addition. This dissociation between difficulty and strategy use occurs because procedural strategies for addition are relatively efficient, and memory strength for multiplication facts is relatively low (Campbell & Xue, 2001). Consequently, if ORP recognition performance is strongly yoked to problem difficulty because of difficulty-related switch costs, then recognition performance should, if anything, be slower following multiplication than following addition, despite more use of procedures for addition.

Finally, in addition to manipulating operation (i.e., addition vs. multiplication) we also manipulated problem size (small if $N \times M \leq 25$, large if $N \times M > 25$) and operand format (Arabic digits vs. written number words). Numerous studies indicate that both of these factors affect use of procedural strategies (see Campbell & Epp, 2005 for a review). Specifically, procedure use increases with problem size, and self-reported use of procedures is much more common with written word operands (four + eight) than with Arabic digits (4 + 8), especially for simple addition (Campbell & Fugelsang, 2001; Campbell, Parker, & Doetzel, 2004). Consequently, we expected verbal-self reports to replicate previous research with respect to the effect of these factors. Given that problem difficulty is greater for large than for small problems, and greater for word relative to digit format, we also expected recognition performance to be relatively poor following large problems relative to small problems, and to be relatively poor following word than digit problems.

Experiment 1

Method

Participants. Forty-eight volunteers (mean age 19 years, 29 female, 44 right handed) were recruited from the introductory psychology participant pool at the University of Saskatchewan. All participants received course credit for their participation.

Apparatus and Stimuli. The experiment was conducted using E-prime software on a PC-type computer. Stimuli were displayed simultaneously to the participant and the experimenter on separate monitors. Arithmetic and recognition answer response times (RT) were collected by voice activated microphone. Arithmetic answers and strategy self-reports were recorded by the experimenter on the keyboard number pad. Recognition answers were collected by button box. Timing was accurate to ±1 ms. Stimulus characters were approximately 4 mm wide × 6 mm high and stimuli were presented at central fixation. Problem stimuli were the 28 single-digit addition or multiplication problems composed of the numbers 2 through 9, excluding tie problems (e.g., 2 + 2, 3 + 3, 4 + 4 etc.). The surface format of the stimuli alternated between digit format (e.g., 3 + 7) and word format (e.g., three + seven). Recognition stimuli were based on problem stimuli and could be true: matching one of the two operands in the preceding problem (3 + 7 e.g., 3 or 7), or false: ±1 number from either operand of preceding problem (e.g., for 3 + 7: 2, 4, 6, or 8). Recognition stimuli were chosen with the further limitation that they could not match the solution to the arithmetic problem.[14]

---

[14] For “2” and “9” problems (e.g., 2 + 3, 3 + 9, etc.), 1 and 10 were not recognition options. For 8 + 9 = 17 the false recognition target was set to 6 to avoid the repetition of the 7 in the solution (i.e., 17).

Design. Arithmetic operation (addition or multiplication) was a between-participants factor. The experiment included four blocks of 56 trials. In each block, the 28 “non-tie” problems composed from pairs of different digits between 2 and 9 were presented once in digit format and once in word format. The order of problems in each block was random with the constraint that digit and word versions of the same problem were separated by at least 12 trials; format alternated across trials. Operand order within blocks was yoked for identical digit and word problems and counterbalanced across blocks: In the first block, half of the problems were randomly selected to be tested with the smaller operand on the left (e.g., 4 × 5 and four × five). Operand order then alternated across blocks. To prevent participants from optimizing performance for operand recognition, half of the four trials in each Problem × Format cell of the design were randomly selected to be followed by a recognition probe, the other half were not. For each problem, one recognition trial was randomly selected to be true, and the other false. Recognition stimuli were yoked such that the same true and false targets were presented for digit and word trials of a given problem. Each recognition target was equally likely to be based on the smallest or largest operand from the preceding problem (e.g., 3 or 7 for 3 + 7) and each number from 2 to 9 was equally likely to appear as a recognition target over the course of the experiment.

Procedure. Participants sat approximately 60 cm from the monitor. A five-key button box with keys 1 and 5 labeled yes or no was positioned on a table directly in front of the participant. A stand suspended a microphone at the appropriate level to detect the onset of verbal response. An experimenter observed from a second monitor located behind and to the left of the testing station and recorded problem answers and strategy reports using the keyboard number pad. Instructions varied by condition. In the addition condition participants received the following instructions:

> You will receive a series of basic addition problems. The format of these problems will alternate between digits (1, 2, etc.) and words (one, two, etc.). When the problem appears on the computer screen, verbally answer the problem as quickly as possible. On half of the trials, after you answer a problem a single digit number (1 or one, 2 or two, etc.) will appear. If it is one of the two addends in the problem that you solved, select the Y key for YES on the button box. If it does not match, select the N key:

7 + 1	7 + 1
say 8	say 8
7	0
select YES	select NO

> After this, a question will appear asking you how you solved the original math problem (7+1):
>
> Transform, Count, Remember, Other?
>
> Please review the solution types before you begin. Before the experiment there will be a Practice Phase.

The experimenter also read through a paper copy of the strategy definitions, which the participant retained throughout the experiment (Campbell & Xue, 2001). The paper copy of the definitions read as follows:

> TRANSFORM: You solve the problem by referring to a related problem in the same or another operation. For example, you might solve 17 - 9 = ? by remembering that 17 -10 = 7, so 17 - 9 must equal 8. COUNT: You solve the problem by counting a certain number of times to get the answer. REMEMBER: You solve the problem by just remembering or knowing the answer directly from memory without any intervening steps. OTHER: You may solve the problem by a strategy unlisted here, or you may be uncertain how you solved the problem.

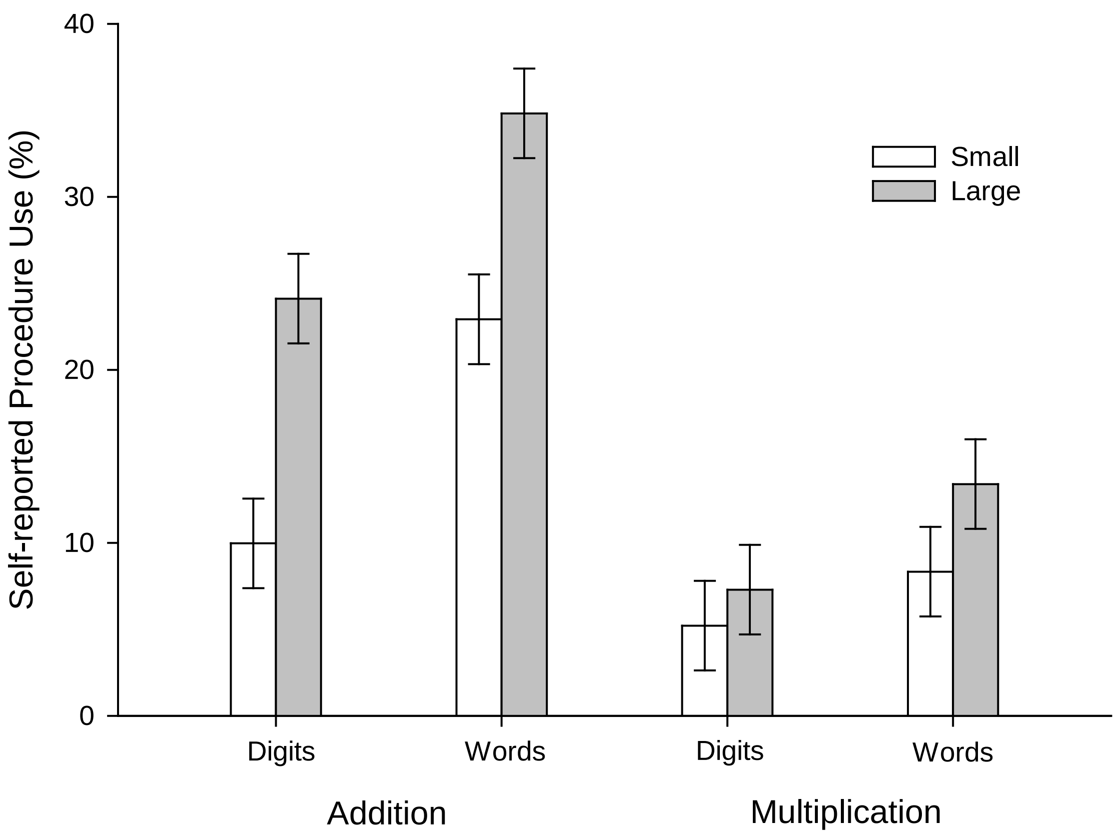
In the multiplication condition participants received identical instructions with reference to addition replaced with reference to multiplication. For each trial, problem stimuli were presented simultaneously (e.g., 3 + 7), preceded by a 500 ms fixation target. Participants verbally reported the solution. On recognition trials, the recognition probe was presented immediately after the arithmetic problem. Participants responded with the button labeled Y or N. For example for the problem 3 + 7, a correct recognition answer would be yes to 3 or 7, and no to 2, 4, 6, or 8. The strategy probe (i.e., Transform, Count, Remember, Other?) immediately followed. On no-recognition trials, the strategy probe immediately followed the report of the arithmetic answer. Participants verbally reported the answer to the strategy probe. The experimenter recorded strategy reports on the keyboard number pad (1=transform, 2=count, 3=remember, 4=other). After this rapid sequence was complete, the experimenter recorded the arithmetic answer reported and noted any failures of the voice activated switch due to extraneous noise or failures to detect onset of response. The next trial commenced.

There was a short break between each block of 56 trials and the duration of testing was approximately 30 minutes. Assignment of left and right keys to yes and no responses was counterbalanced by participant. Participants were familiarized with the trial event sequence in a practice phase that included eight 0 and 1 problems (e.g., 0 + 7, 1 + 9) not included in the experimental stimuli. Each practice problem included a recognition probe.

Results

Strategy Reports. Strategy self-reports following incorrect answers to arithmetic (5.1%) or recognition (3.1%) were excluded. A further 0.3% of reports were not recorded due to experimenter error. Non-retrieval strategies (i.e., transform, count, other) were summed to produce an index of total procedure use. Problem size for both addition and multiplication was defined as “small” if the product of the two operands was ≤ 25 and “large” if the product was > 25 (cf. Campbell, Kanz & Xue, 1999; Campbell & Xue, 2001). Figure 5-1 presents mean percent of total procedure use and Table 5-1 presents mean percent of retrieval use and the individual procedure types as a function of operation (addition or multiplication), format (digits or words), and problem size (small or large). The respective 2 (between) × 2 × 2 (within) mixed-measures analysis of variance (ANOVA) was conducted on total procedure use. The ANOVA is summarized in Table 5-2.

Figure 5-1. Mean Percentage Reported Procedure Use as a Function of Operation, Problem Size, and Format for Experiment 1



Note. Error bars are 95% mixed-measures confidence intervals (Jarmasz & Hollands, 2009).

The results for all main effects were as expected. Reported procedure use was higher for addition problems (23%) than for multiplication (9%). Digit format problems were reportedly solved with fewer procedures (12%) than problems in written word format (20%). Small problems were solved with fewer procedures (12%) than large problems (20%). These main effects were qualified by the Operation × Size and Operation × Format interactions. For the Operation × Size interaction, the problem size effect (large – small) was larger for addition problems (+13% procedures) than for multiplication problems (+4%) (cf. Campbell & Xue, 2001). For the Operation × Format interaction, the word-format cost (word – digits) was larger for addition (+12% procedures) compared to multiplication (+5%) (cf. Campbell et al., 2004). The Format × Size and the Operation × Format × Size interactions were not significant.[15] The strategy report data largely replicated previous results; therefore, there is no concern that the recognition procedure fundamentally altered participants' strategy reports.

Table 5-1. *Mean Percentage Reported Use of Solution Strategies as a Function of Operation, Format, and Problem Size for Experiment 1*

	Addition			Multiplication		
Strategy	Small	Large	All	Small	Large	All
Digit Format						
Remember	87	65	76	91	81	86
Count	8	12	10	5	4	5
Transform	2	12	7	0	3	2
Other	0	0	0	0	0	0
Word Format						
Remember	71	49	60	85	71	78
Count	19	17	18	8	7	8
Transform	4	17	11	0	5	3
Other	0	0	0	0	1	1

Note. Remember = directly retrieved the answer from memory. Transform = referred to another related problem. Count = counted by ones to get the answer. Other = used a strategy unlisted here or uncertain.

[15] The typical Format × Size interaction (i.e., greater word format costs on procedure use for large problems; e.g., Campbell & Alberts, 2009; Campbell & Fugelsang, 2001) approached significance ($p = .07$) in an analysis that included all trials. For the main analysis, however, we included only strategy self-reports that followed correct answers in order to match trials used for the strategy analysis with those used for the analysis of recognition RT.

Table 5-2. *Operation × Format × Problem Size Analyses of Variance for Mean Percentage of Procedure Use, Response Time, and Error Percentage for Arithmetic, and Mean Response Time and Error Percentage for Recognition (Experiment 1)*

	F				
	Arithmetic			Recognition	
Source	% procedure	RT	% error	RT	% error
Operation (O)	10.7***	0.0	2.9+	0.0	2.2
MSE	933.1	2311587.3	60.0	231898.6	1.8
Format (F)	24.7***	316.8***	12.6***	47.7***	11.1***
MSE	131.5	23098.4	14.8	6559.6	0.7
Size (S)	13.2***	89.2***	52.3***	72.9***	8.4**
MSE	250.9	48505.2	35.9	5324.9	1.2
O × F	4.8*	0.0	0.1	0.0	1.1
MSE	131.5	23098.4	14.8	6559.6	0.7
O × S	4.3*	0.0	2.8	1.6	3.0+
MSE	250.9	48505.2	35.9	5324.9	1.2
F × S	0.0	29.1***	1.5	8.8**	1.2
MSE	39.7	4867.2	9.7	3430.3	1.1
O × F × S	2.0	0.0	0.2	1.4	0.2
MSE	39.7	4867.2	9.7	3430.3	1.1

Note. % procedure = mean percentage of combined non-retrieval strategies reported. RT = correct mean median response time. % error = mean error percentage. Recognition RT and % error exclude trials following incorrect arithmetic answers. For all tests, 1 and 46 degrees of freedom. $^{+}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .005$.

Arithmetic Performance. Median correct arithmetic RT and percent errors received parallel ANOVAs. The results of these analyses appear in Table 5-2 and the means appear in Table 5-3. Response time data for 2.8% of trials were discarded due to failure of the microphone to detect the onset of response.

Table 5-3. *Mean Median Response Times and Error Percentage for Arithmetic and Recognition as a Function of Operation, Format, and Problem Size for Experiment 1*

	Addition			Multiplication		
Measure	Small	Large	L-S	Small	Large	L-S
Digit Format						
Arith RT	780	1028	248	774	1018	244
(% error)	1.0	5.1	4.1	1.4	8.8	7.4
Recog RT	1022	1063	41	996	1085	89
(% error)	0.5	1.0	0.5	0.5	0.6	0.1
Word Format						
Arith RT	1112	1471	359	1114	1464	350
(% error)	2.4	8.0	5.6	2.9	10.9	8.0
Recog RT	1067	1179	112	1063	1181	118
(% error)	0.8	1.8	1.0	0.7	1.0	0.3

Note. Arith RT = correct mean median arithmetic response time. Recog RT = correct mean median recognition response time. % error = mean error percentage. Recognition RT and % error exclude trials following incorrect arithmetic answers. Response times in milliseconds.

The expected main effects of format and problem size were observed. Digit problems were solved faster (900 ms) than word problems (1290 ms) and small problems were solved faster (945 ms) than large (1245 ms). Also as expected for the Format × Problem size interaction, the problem size effect was larger for problems presented in word format (+355 ms) compared to digit format (+246 ms) (cf. Campbell, 1994). The analysis indicated no RT differences between addition (1098 ms) and multiplication (1092 ms), $F < 1$, and there was no evidence of two-way or three-way interactions of operation with format or problem size (all *F's* < 1). As we discuss further on, this contrasts with the results of several previous studies (e.g., Hecht, 1999; Campbell & Arbuthnott, 2010) that manipulated operation (i.e., addition vs. multiplication) as a within-participants factor.

Consistent with the RT analysis, the error analysis indicated more errors for word format (6.0%) than for digit problems (4.1%) and more errors for large (8.2%) than small problems (1.9%). There were no other significant effects in the error analysis although there was a non-significant trend for more errors on multiplication problems (6.0%) than for addition problems (4.1%), $p = .10$.

Recognition Performance. Median correct recognition RT and percent errors (Table 5-3) were analyzed in separate ANOVAs (Table 5-2). Recognition RTs on trials with an arithmetic error (5.1%) were not included in the analysis. Consistent with the results of Thevenot et al. (2007), recognition was faster following small problems (1037 ms) compared to large (1127 ms). Recognition also was faster following digit format problems (1042 ms) compared to word format problems (1122 ms). These two main effects were qualified by the Format × Size interaction: The problem size effect on recognition RT was larger following problems in word format (+115 ms) compared to digit format (+65 ms). Importantly, there was no difference in recognition RT following addition (1083 ms) compared to following multiplication (1081 ms), $F < 1$, and operation did not participate in any interaction effects (all $p \geq .21$). Thus, whereas there was a large effect of operation on reported procedure use there was no evidence for a corresponding effect on recognition response time. This indicates that strategy self-reports and the ORP diverge in their implications for strategy differences between addition and multiplication.

The overall rate of recognition errors was very low (0.9%). There were more recognition errors following problems in word format (1.1%) compared to digit (0.7%) and large problems (1.1%) compared to small (0.6%). There was a trend for a larger problem size effect for recognition following addition problems (+0.8%) compared to multiplication (+0.2%), but the test for the interaction only approached significance, $p = .09$. All other interactions were not significant (all $p \geq .27$).

Discussion

The purpose of Experiment 1 was to assess the degree to which strategy self-reports and ORP performance agree in their implications about strategy use for simple addition and multiplication. The agreement was only partial. The main effects of format and size were consistent for both measures: Relative increases in procedure reports from digits to words and small to large problems were matched by corresponding increases in recognition RT. This would be expected if recognition RT and strategy reports both detect variability in use of procedural strategies as a function of these factors. In contrast, however, the effects of operation were inconsistent for reported procedure use and recognition RT. The strategy reports indicated far more use of procedures for addition (23%) than multiplication (9%), which is consistent with previous research (e.g., Campbell & Xue, 2001; LeFevre, et al., 1996a); but there was no effect of operation on recognition RT. As there was also no effect of operation on arithmetic RT or errors, the pattern of results is consistent with the view that recognition performance was sensitive to the difficulty (i.e., RT) of the immediately preceding arithmetic problem rather than strategy per se. Metcalfe and Campbell (2010) provided evidence of difficulty-related switch costs in the ORP. Indeed, here, the pattern of significant effects on arithmetic RT exactly matched that of recognition RT (see Table 5-2).

Nonetheless, with respect to strategy use, the point of disagreement between the strategy reports and ORP arose from the operation factor, and there were potentially atypical effects of operation in Experiment 1. Specifically, unlike in previous research, multiplication was not more difficult (i.e., slower and more error prone) than addition (e.g., Campbell, 1994; Campbell & Arbuthnott, 2010; Campbell & Xue, 2001). For example, Campbell and Arbuthnott found multiplication to be slower overall and with a larger problem size effect relative to addition. A key difference between Experiment 1 and previous research comparing addition and multiplication is that operation was a between-participants factor here whereas it was within-participants in comparable previous experiments. The relative difficulty of multiplication compared to addition may be more difficult to detect in a between-participants design. Furthermore, the fact that strategy reports and recognition RT agreed with respect to the effects of the within-participants factors (size and format), but not the between-participant factor (operation), raises the possibility that experimental design contributed to this discrepancy. Consequently, in Experiment 2, we repeated Experiment 1 but treated operation as a within-participants factor. Based on previous research, we expected multiplication to be more difficult than addition. If ORP performance is controlled by problem difficulty owing to switch costs then, unlike Experiment 1, recognition RT should if anything, be slower following multiplication than addition in Experiment 2.

Experiment 2

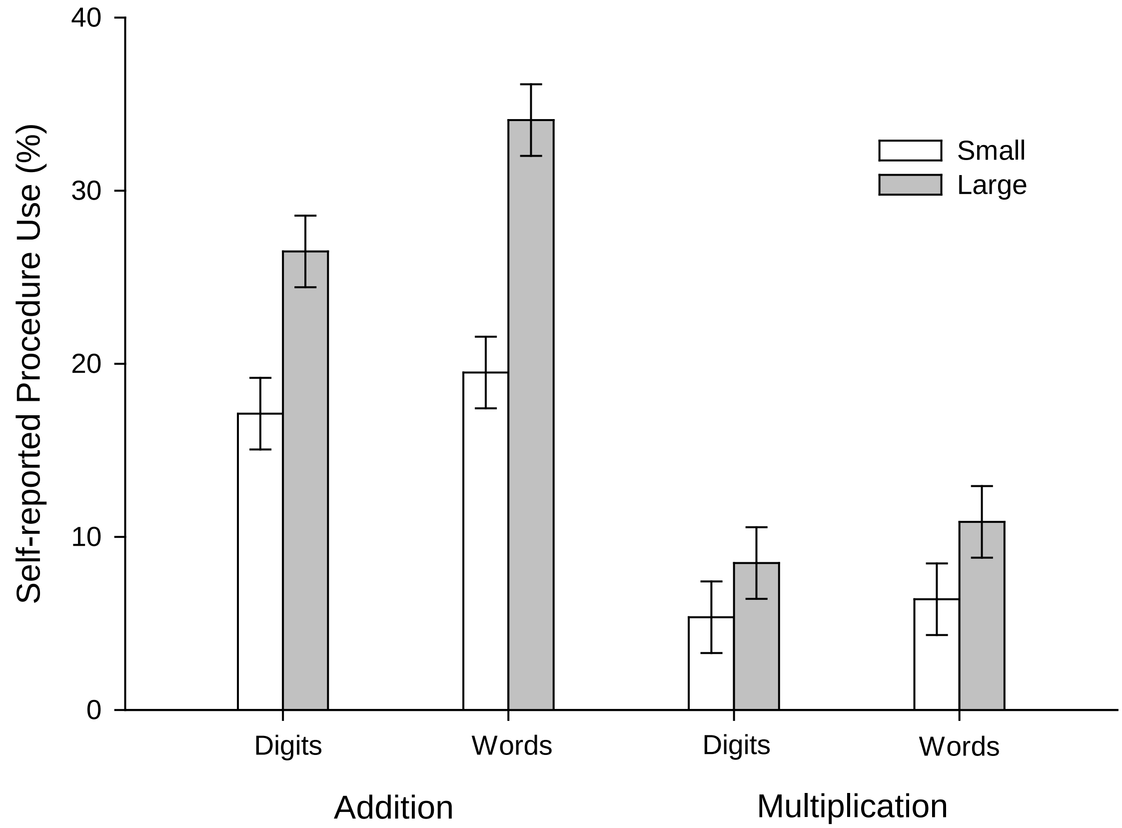
Method

Twenty-four volunteers (mean age 26 years, 15 female, 21 right handed) were recruited using posters advertising a psychology study at the University of Saskatchewan. Each participant received $10 for their participation. The method for Experiment 2 was based on the method used for Experiment 1. The major exception was that in Experiment 2, arithmetic operation was treated as a within-participants factor. Each participant was presented with four blocks of addition and four blocks of multiplication for a total of eight blocks of 56 trials. Operation alternated across blocks and the operation of the first block was counterbalanced. To ensure that the arithmetic response was completed before the display of the recognition probe a 250 ms SRI was inserted between the onset of the answer and the onset of recognition stimuli. Instructions from Experiment 1 were modified to describe addition and multiplication trials. Practice problems from Experiment 1 were used in Experiment 2 with the modification that a random half of the problems were presented as addition problems and half were presented as multiplication problems. The duration of testing was approximately 50 minutes.

Results

Strategy Reports. Strategy self-reports following arithmetic errors (3.4%) or recognition errors (2.1%) were excluded from analysis. The mean percentage of total procedure use (Figure 5-2) and the mean percentage of individual strategies used (Table 5-4) are presented as a function of operation (addition or multiplication), format (digits or words), and problem size (small or large). Table 5-5 presents the results of the repeated measures ANOVA for total procedure use (i.e., Figure 5-2). The pattern of procedure rates for Experiment 2 was similar to Experiment 1. Reported procedure use was higher for addition problems (24%) than for multiplication (8%). Digit format problems were reportedly solved with fewer procedures (14%) than problems in written word format (18%). Small problems were solved with fewer procedures (12%) than large problems (20%). Consistent with previous research (e.g., Campbell & Fugelsang, 2001), there was a Format × Size interaction in Experiment 2; the observed problem size effect (large – small) for words was larger than for digits (10% and 6%, respectively). As in Experiment 1, the Operation × Format × Size interaction was not significant.

Figure 5-2. Mean Percentage Reported Procedure Use as a Function of Operation, Problem Size, and Format for Experiment 2



Note. Error bars are 95% mixed-measures confidence intervals (Jarmasz & Hollands, 2009).

Table 5-4. *Mean Percentage Reported Use of Solution Strategies as a Function of Operation, Format, and Problem Size for Experiment 2*

	Addition			Multiplication		
Strategy	Small	Large	All	Small	Large	All
Digit Format						
Remember	80	68	74	93	83	88
Count	15	7	11	5	1	3
Transform	2	19	11	0	6	3
Other	0	1	1	0	2	1
Word Format						
Remember	76	59	68	90	78	84
Count	17	10	14	5	2	4
Transform	3	24	14	2	7	5
Other	0	1	1	0	3	2

Note. Remember = directly retrieved the answer from memory. Transform = referred to another related problem. Count = counted by ones to get the answer. Other = used a strategy unlisted here or uncertain.

Table 5-5. *Operation × Format × Problem Size Analyses of Variance for Mean Percentage of Procedure Use, Response Time, and Error Percentage for Arithmetic, and Mean Response Time and Error Percentage for Recognition (Experiment 2)*

	F				
	Arithmetic			Recognition	
Source	% procedure	RT	% error	RT	% error
Operation (O)	9.8**	10.6***	12.1***	0.2	6.1*
MSE	1340.1	124900.4	10.6	12450.1	5.8
Format (F)	13.8***	114.1***	4.2+	58.0***	5.2*
MSE	38.9	75390.4	22.8	5188.7	10.7
Size (S)	4.8*	49.0***	21.9***	46.6***	3.3+
MSE	628.1	126865.2	56.9	7270.7	14.6
O × F	7.6*	0.0	0.1	1.2	1.5
MSE	17.0	11795.7	5.5	4852.3	5.3
O × S	7.8*	2.8	17.2***	0.1	0.3
MSE	103.3	72287.6	10.5	4856.3	9.6
F × S	10.5***	10.2***	1.0	2.7	2.0
MSE	12.3	8095.4	6.3	3956.7	10.0
O × F × S	1.4	4.6*	3.1+	0.0	0.2
MSE	32.3	8233.2	3.6	5557.7	2.5

Note. % procedure = mean percentage of combined non-retrieval strategies reported. RT = correct mean median response time. % error = mean error percentage. Recognition RT and % error exclude trials following incorrect arithmetic answers. For all tests, 1 and 23 degrees of freedom. $^{+}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .005$.

Arithmetic Performance. Previous research manipulating operation (i.e., simple addition vs. multiplication) as a within-participants factor found multiplication to be slightly more difficult than addition (e.g., Hecht, 1999). Median correct arithmetic RT and percent of errors were analyzed using separate 2 (Operation) × 2 (Format) × 2 (Size) repeated-measures

ANOVAs. RT data for 1.3% of trials were discarded due to microphone failures. The results of these analyses appear in Table 5-5 and the means appear in Table 5-6.

As expected, there was a main effect of operation: Multiplication problems (1372 ms) were answered more slowly than addition (1206 ms), and all cells in the design contributed to this overall effect (see Table 5-6). There was also an Operation × Format × Size interaction. The form of the 3 way interaction was the same as that observed by Campbell (1994): the problem-size effect for addition was much larger for words (+364 ms) than digits (+225 ms, $p \leq .001$), whereas the problem size effect for multiplication with words and digits were similar (+438 ms and +414 ms, respectively, $F < 1$). The main effects of format and size, and the Format × Size interaction were all significant and of the same form as in Experiment 1.

The error ANOVA indicated fewer errors for addition (4.1%) than multiplication (5.7%). There was a trend ($p = .052$) for less errors with digit (4.2%) compared to word format (5.6%), and there were fewer errors for small problems (2.4%) compared to large (7.5%). In contrast to Experiment 1, the Operation × Size interaction was significant: the problem-size effect on errors was larger for multiplication problems (+7.0%) than addition (+3.2%). All other interactions were non-significant ($p \geq .092$). The effects of the within-participants manipulation of operation in Experiment 2 are in clear contrast to the null effects of operation observed in Experiment 1 and indicate an effect of treatment type (between vs. within) on the relative difficulty of addition and multiplication as indexed by arithmetic RT and error rate.

Table 5-6. *Mean Median Response Times and Error Percentage for Arithmetic and Recognition as a Function of Operation, Format, and Problem Size for Experiment 2*

	Addition			Multiplication		
Measure	Small	Large	L-S	Small	Large	L-S
Digit Format						
Arith RT	883	1108	225	954	1366	412
(% error)	1.7	5.0	3.3	2.0	8.2	6.2
Recog RT	952	1023	71	971	1039	68
(% error)	1.8	2.5	0.7	0.9	0.9	0.0
Word Format						
Arith RT	1235	1599	364	1366	1804	438
(% error)	3.3	6.4	3.1	2.5	10.3	7.8
Recog RT	1025	1129	104	1026	1120	94
(% error)	1.9	3.7	1.8	1.6	3.1	1.5

Note. Arith RT = correct mean median arithmetic response time. Recog RT = correct mean median recognition response time. % error = mean error percentage. Recognition RT and % error exclude trials following incorrect arithmetic answers. Response times in milliseconds.

Recognition Performance. Median correct recognition RT and percent recognition errors (Table 5-6) received corresponding ANOVAs (Table 5-5). Recognition trials following arithmetic errors (3.4%) were not included in the analysis.

As in Experiment 1, recognition RT was longer following word format problems (1075 ms) compared to digit format problems (996 ms), and longer following large (1078 ms) compared to small problems (994 ms). The form of the Format × Size interaction was similar to Experiment 1, with size effects tending to be larger for words (+99 ms) than digits (+71 ms); but unlike Experiment 1 the interaction was not significant (p = .12). As in Experiment 1, there was no main effect of operation on recognition RT (1032 ms following addition and 1039 ms following multiplication, $F < 1$), and operation did not participate in any interaction effect on

RT. Thus, recognition RT did not vary with operation despite a three-fold higher rate of self-reported procedure use for addition (24%) than multiplication (8%).

The overall rate of recognition errors was low (2.1%). The ANOVA indicated only main effects (see Table 5-5). There were more recognition errors following word (2.6%) compared to digit problems (1.5%), and a trend for more errors following large (2.6%) than small problems (1.6%, p = .082). Finally, there were more recognition errors following addition than (2.5%) multiplication (1.6%). These main effects correspond to parallel effects on reported strategy use and therefore, with respect to these main effects, the ORP and self-reported strategies agree.

Separating Effects of Strategy and Problem Difficulty on Recognition Performance

Whereas strategy reports and ORP recognition performance agreed with respect to main effects of format and size, the strategy reports, but not the ORP, implied more procedure use for addition than multiplication. The null effect of operation on recognition RT might reflect opposing influences of strategy use and relative problem difficulty across operations. That is, the relative difficulty of multiplication would tend to increase recognition RT relative to addition (i.e., owing to difficulty-related switch costs), while more use of procedures for addition than multiplication would tend to slow addition recognition RT relative to multiplication (i.e., owing to procedure-related degradation of operand representations). In this way, effects of problem difficulty (i.e., switch costs) and strategy (i.e., operand degradation) might be confounded. Using subsets of the data, however, it is possible to disconfound potential effects of problem difficulty and strategy. To this end, we collapsed the data over format and computed mean correct RT for small addition problem *procedures* (1519 ms) and large multiplication *retrieval* trials (1607 ms) for the 15 participants who contributed observations to both categories of trials. These two cells are matched for difficulty, but recognition RT was longer following reported procedures (1286 ms) compared to retrieval (1102 ms), $t(14) = 2.7$, $SE = 68.4$, $p = .018$. Similarly, we compared recognition performance following procedures reported for large addition problems in digit format to recognition following reported retrieval for large multiplication in digit format. These arithmetic cells are closely matched for difficulty among the 17 participants who contributed observations to them (means of 1475 ms and 1402 ms for procedures and retrieval, respectively). Again, although the cells are matched for arithmetic difficulty, recognition RT was longer following reported procedures (1243 ms) compared to retrieval (1113 ms), $t(16) = 2.1$, SE = 60.2, p = .048. These analyses reinforce the conclusion that recognition RT in the ORP is sensitive to strategy use for simple arithmetic when problem difficulty is controlled.[16]

On the other hand, it is also possible to use a subset of the data to hold arithmetic strategy constant and examine recognition performance as a function of problem difficulty. Accordingly, we recalculated the means in Table 5-3 (Experiment 1) and Table 5-6 (Experiment 2) after excluding all self-reported procedure trials. Thus, the recalculated means included only reported retrieval trials and arguably represented use of a single, uniform retrieval strategy. Nonetheless, across the eight experimental cells, mean arithmetic retrieval RT was positively correlated with mean recognition RT in both Experiment 1 [$r(6) = +.78$, $b = .131$, $p = .02$] and Experiment 2 [$r(6) = +.92$, $b = .208$, $p = .001$]. The corresponding correlations between mean arithmetic error rate and mean recognition RT was +.46 ($p = .25$) in Experiment 1 and +.75 ($p = .03$) in Experiment 2. Thus, with strategy variability controlled, mean problem difficulty across the eight cells of the experiment provided good prediction of recognition RT in both experiments.

The preceding analyses suggest that both procedure use and problem difficulty affected recognition RT. To estimate the relative effect size of these influences we performed multiple regressions to predict mean recognition RT as a function of problem difficulty and procedure use across the combined 16 cells from Experiment 1 (Table 5-3) and Experiment 2 (Table 5-6). Mean recognition RT was correlated .718 and .788 with arithmetic RT and error rate, respectively (both $p \leq .002$). The zero-order correlation of recognition RT and procedure use (r = .440) did not reach significance ($p = .088$). Owing to the high correlation between arithmetic RT and arithmetic error rate [$r(14) = .731$, $p = .001$] we performed separate multiple regressions on recognition RT with, respectively, arithmetic RT and procedure rate, and arithmetic error and procedure rate as predictors. The standardized beta weights for the two predictors in each analysis and the significance tests for the overall models appear in Table 5-7. As Table 5-7 shows, the regression model including arithmetic RT and procedure use was significant (adjusted $R^2 = .496$). With arithmetic RT and procedure rate entered simultaneously into the regression,

---

[16] We do not report the corresponding analyses for Experiment 1 because we could not equate procedure and retrieval trials for arithmetic difficulty.

arithmetic RT was a robust predictor of recognition RT across cells, whereas procedure use was not a significant predictor of recognition RT (p = .254). The analysis with mean arithmetic error rate and procedure use as predictors yielded an adjusted R^2 of .663. Again, arithmetic performance or difficulty, in this case arithmetic error rate, was a good predictor of recognition RT across cells. Procedure use, on the other hand, was only a marginally significant predictor of recognition RT (p = .07). The non-significant beta weights for procedure reports in both the RT and error rate regressions indicates that problem difficulty was more predictive of recognition RT than were mean procedure reports.

Table 5-7. *Standardized Beta Weights from Regression of Mean Recognition RT on Arithmetic Performance (RT or Errors) and Procedure Use*

RT and Procedures		Errors and Procedures	
RT	.64**	Errors	.73***
Procedures	.23	Procedures	$.30^{+}$
$R = .75$, $F = 8.4$, $p = .005$		$R = .84$, $F = 15.8$, $p < .001$	

Note. $^{+}p < .1$, $^{**}p < .01$, $^{***}p < .005$. Degrees of freedom were $F(2, 13)$ for the F test of the models.

General Discussion

With respect to the factors of problem size and format, Experiment 2 largely reproduced the same pattern of effects as in Experiment 1 for all five dependent measures (compare Tables 5-2 and 5-5). The experiments differed, however, with respect to effects of the operation manipulation. Unlike Experiment 1, multiplication in Experiment 2 was more difficult than addition and produced a larger problem-size effect than addition for both RT and errors. Testing operation as a within-participants variable makes these performance differences more readily apparent compared to between-participants manipulation of operation. We hypothesized that if recognition performance is sensitive to the difficulty of the immediately preceding problem, then the effects of operation (i.e., multiplication more difficult than addition) would be reflected in relatively reduced recognition performance following multiplication compared to following addition. In fact, recognition errors were slightly more likely following addition than multiplication, but arithmetic operation had no effect on mean recognition RT. As we would expect more recognition errors following procedures than retrieval (i.e., Thevenot et al., 2007), and procedural strategies were reported far more often for addition than multiplication, the recognition error data are potentially consistent with the theory that the ORP is sensitive to strategy choice.

Indeed with respect to the main effects of operation, format, and problem size, recognition error rates and strategy self-reports were in good agreement (i.e., increased procedure use was associated with more recognition errors). Furthermore, recognition RT reflected corresponding effects of size and format on procedure use. With respect to these variables the ORP and strategy self-reports also agreed. Nonetheless, neither recognition RT nor recognition errors presented the Operation × Format nor Operation × Size effects observed in the strategy reports in both experiments or the Format × Size effect in Experiment 2. Thus, if the strategy reports are considered to be generally valid, then the verbal self-reports provided a much more sensitive measure of interaction effects than did ORP recognition performance.

Our results suggest that recognition performance in the ORP is sensitive both to arithmetic strategy (i.e., procedures vs. direct retrieval) and difficulty-related switch costs. We presented multiple regression analyses to estimate the relative effect size of these factors. Using the strategy self-reports to estimate procedure use, we found that the rate of procedure use was a marginally significant predictor of recognition RT across the 16 cells of the two experiments, but that arithmetic difficulty (measured as mean RT and error rate) were more successful predictors of recognition RT. Indeed, the beta weights for arithmetic difficulty were relatively large compared to those for procedure use. Thus, with respect to simple addition and multiplication, the regression results suggest that recognition RT in the ORP is influenced at least as much by the relative difficulty of the preceding problem (presumably owing to switch costs) than by the strategy used to solve that problem (i.e., procedures vs. direct memory retrieval). One might argue that the relatively high beta weights for arithmetic difficulty reflect a confound between problem difficulty and strategy use, because procedural strategies are generally slower and error prone relative to direct retrieval. Contrary to this, however, mean reported procedure use across the 16 cells of the two experiments was not significantly correlated with either mean RT (.324, *p*

= .22) or error rate (.189, $p = .48$). This suggests that, in these experiments, overall problem difficulty and mean rate of procedure use were relatively independent.

In Experiment 2, multiplication was more difficult than addition with respect to both RT and errors. Given our argument that recognition RT increases with the difficulty of the preceding problem, how can the null effect of operation on recognition RT be explained? First, opposing influences of strategy use and relative problem difficulty across operations would work against an effect of operation on recognition performance. Although multiplication was more difficult than addition in Experiment 2, the rate of procedure use was 68% higher for addition than multiplication. Thus, operand degradation owing to addition procedure use would obscure difficulty-related switch-costs owing to the greater difficulty of multiplication. Furthermore, the overall effect of operation on arithmetic difficulty (+166 ms, $\eta^2 = .32$) was quite small compared to either format (+423 ms, $\eta^2 = .83$) or problem size (+360 ms, $\eta^2 = .68$). Thus, we would expect the effect of operation difficulty on recognition RT to be relatively small. In contrast, the absence of an operation effect on ORP performance in Experiment 1, despite evidence from the strategy reports that procedures were used far more often for addition than multiplication, suggests that the ORP may be insensitive to factors manipulated between participants.

Conclusions

These experiments were the first to test recognition performance in the ORP using the complete standard set of basic addition and multiplication problems (i.e., non-tie problems composed from operands 2 to 9), and the first to test recognition performance in a production task instead of a verification task. The results indicated partial agreement between operand recognition performance and verbal self-reports of strategy use for simple addition and multiplication; specifically, both measures agreed with respect to main effects of problem size and operand format on procedure use, but the strategy reports indicate interaction effects not revealed by ORP performance. The results also indicated, furthermore, that ORP performance was affected by the difficulty of the preceding problem, presumably owing to difficulty-related switch costs (Metcalfe & Campbell, 2010). Finally, despite a variety of evidence, both from the current experiments and others (e.g., Campbell & Xue, 2001), that simple addition relies much more on procedural strategies than multiplication, ORP recognition performance showed little or no sensitivity to operation. This was true both when operation was manipulated between-participants in Experiment 1 and when operation was manipulated within-participants in Experiment 2. This null result appears to represent a basic failure of the paradigm to detect a robust and widely recognized effect of operation (i.e., addition vs. multiplication) on procedure use. Thus, while the validity and precision of verbal self-reports must continue to be evaluated cautiously (e.g., Kirk & Ashcraft, 2001; Thevenot et al., in press), the present findings indicate that the ORP is not a reliable substitute for verbal reports to measure strategy choices for simple arithmetic.

CHAPTER 6

General Conclusions

The primary thesis of this dissertation was to increase knowledge about strategy use for basic arithmetic. Chapter 2 investigated the role of operand familiarity in adults' performance of simple addition. Two experiments provided evidence that pre-exposure to single-digit operands increased reported use of direct retrieval for new combinations of the familiarized operands. RT and error patterns across experiments also supported the conclusion that increased use of retrieval facilitated performance. The results extend the domain of application of cue familiarity effects (cf. Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992; Schunn et al., 1997) to include elementary arithmetic and accordingly, models of strategy choice for arithmetic should consider the potentially important role of operand familiarity.

In the expanded context of arithmetic cognition as a whole, the effect of written word format on arithmetic has been the basis for proving solution processes are partially connected to modal stimuli as opposed to earlier theories that viewed solution processes as unified mental abstractions. Despite the ubiquity of written word format as an experimental manipulation there was no consensus on why it produced its hallmark reduction in direct memory retrieval when compared to problems presented in the more conventional digit format. Thus the work contained in Chapter 2 was important in a wider sense because it demonstrated that increasing the familiarity of the written word stimuli relative to unrelated neighbours increased reported retrieval rates of relatively well memorized basic addition facts, thus reversing some of the effect of written word format on calculation. This allowed for an expanded understanding for one of the canon findings in the field.

Chapter 3 was the first study in the field to specifically investigate the effect of verbal-auditory arithmetic on strategy choice and only the second study of verbally presented basic arithmetic to date. Filling this void in the literature is important, per se, because of the ubiquity of verbal arithmetic in learning and recall from grade-school to adulthood. Of theoretical importance, the work demonstrated that format affected the solution stage of cognition instead of only basic stimuli encoding processes; this means that models of math knowledge must include a role for specified verbal-auditory solution paths. This result also had important interdisciplinary implications because it provided evidence that behavioural response patterns in members of the normal adult population conform to predictions that were generated based on the neuroimaging and case study literature.

The title of this dissertation is Strategy Use and Basic Arithmetic Cognition in Adults. The insightful reader will note that all of the work contained herein is on the operations of addition and multiplication. This is a common theme in the field of arithmetic cognition, that is, work with the operations of addition and multiplication has historically far outnumbered those concerned with the other main arithmetic operations, subtraction and division. Thus, when addressing the questions in Chapters 2, 4, and 5 it was advantageous to select the operations that had the widest body of background literature associated with them. Furthermore, the stimuli set for addition and multiplication lends itself much more easily to statistical contrasts that can be exactly matched for item effects. The questions addressed in Chapters 2, 4, and 5, however, should be equally applicable to subtraction and division as well, because no major part of these research questions depended on operation specific attributes in their main thesis, other than to make basic predictions used to test other relevant assumptions. This, however, is not the case when we consider the thesis of Chapter 3. Indeed, current literature in the field of cognitive neuroscience suggests that subtraction and division would behave differently than either addition or multiplication as a function of the relative paucity of practice and subsequent retrieval likelihood associated with these respective bodies of arithmetic facts. Future work in this area could provide additional converging data supporting my interpretation of the μ and τ distributions argued in the discussion of Chapter 3. Specifically, while multiplication is touted to be a clear case where a majority of the answers will be directly retrieved, addition represents a hybrid fact set where the smaller facts (e.g., operands < 5, sums < 10, sum^2, products < 25, etc.) are much more likely to be retrieved, whereas the likelihood of larger facts being solved by procedures is much higher. In contrast, procedure rates for subtraction and division are consistently lower than multiplication and addition. Consequently, a direct comparison of either of the former operations to multiplication could be interpreted as likely to produce clearer results vis-à-vis demonstrating a behavioural dissociation between directly retrieved versus computed answers, and an interaction with these strategies and the factor, arithmetic operation.

Much research has used strategy variability to frame theoretical inquiry into the nature of arithmetic cognition in adults. Classically, both chronometric data (e.g. Chapter 3) and metacognitive judgments (e.g. Chapter 2) have been used to infer what type of strategy was used by a participant. Thus many important theories hinge on these data. However, these measures have been criticized as imprecise at best, and completely subject to reactivity on the part of the participant at worst. Therefore the prospect of a new method to measure strategy choice was very important (Thevenot, et al., 2007). However, the findings of Chapter 4 indicate Thevenot et al.'s optimism was unwarranted. Their method of inferring procedure use from operand recognition was found to be confounded with the relative difficulty of the problem, resulting in asymmetrical switch costs masquerading as effects of elaborative processing. At best, this means the new paradigm is as qualified as earlier methods of measurement, at worst these findings combined with the new task demands introduced by the paradigm, render it irrelevant.

One notable addition to the conclusions of Chapter 4 that was not considered at the time of its original publication is that it is possible that the limits of the effects of task switching were understated in the original conclusions. In Chapter 4, it was argued that the poor performance on recognition of the second operand contrasted with the first was evidence that some strategy effects were definitely present in the measure. Implied in this argument is the idea that task switching between cognitive sets would only result in proactive interference effects on recognition task performance (i.e., interference feeding forward affecting initiation of the new task set). Upon further consideration, however, it seems equally likely that the cognitive cost of task switching could result in some form of retroactive interference. That is, it is possible that the cognitive cost of switching tasks could feed backward and interfere with short-to-long-term memory consolidation for the second operand due to the relative shorter interval between presentation of the second operand and the task switch compared to the longer duration between the first operand and the task switch. The precedent for this in the literature would be the primacy effect in the serial position curve. A follow up experiment might test this idea. However, combined with the data presented in Chapter 5, it seems unlikely now that the operand recognition paradigm will be widely adopted in the area. Consequently, I raise this issue here with the expectation that this would be the final opportunity for such discussion.

Chapter 5 was undertaken to address the mixed evidence with respect to the validity and precision of the operand recognition paradigm that was extant at the time the work began. To generate data on the utility of the paradigm for wider application in arithmetic cognition research, the paradigm was modified for use with an answer production task, in contrast to the verification task, as the former is at least as common in the literature as the later, and certainly is higher in ecological validity. For similar reasons, the current work experimented with the complete standard set of addition and multiplication problems from 2 to 9, in contrast to the previous work that used smaller samples of problem sets. Operand recognition performance was directly compared to verbal self-reports to determine the degree of agreement between the two measures. A secondary concern in this work was to assess if the recognition method was worth modifying or correcting in as much as it could replicate predictions for strategy use drawn from the literature on the effects of operation, format, and problem size manipulations on strategy, despite the known issues with problem difficulty. In the key experiment, Experiment 2, results indicated that the operand recognition paradigm was especially ineffective at measuring strategy use differences between addition and multiplication. Whereas the strategy reports detected large differences in strategy use between operations (i.e., more procedures for addition), the recognition data indicated absolutely no trend (not a scintilla, $F < 1$). Thus, the major contribution of this work is that it establishes that the operand recognition paradigm currently lacks sensitivity for use in mainstream arithmetic cognition research.

A secondary finding of the work in Chapter 5 that was not originally predicted, yet may have wide reaching implications for experimental research in the area, was the effect of operation on arithmetic RT in Experiment 1 versus Experiment 2. Specifically, in Experiment 1, for reasons which essentially amount to a random decision by the experimenter, the factor of operation was treated as a between participants manipulation. This resulted in an unexpected null effect of operation that has not been previously described in the literature. Instead, in the literature, operation has routinely been examined between experiments or as a within participants factor. When tested as a within participants factor the results have consistently demonstrated an effect of operation wherein multiplication is slower than addition (albeit a small effect when compared to those of format and size), which has generally been interpreted as proof that multiplication is more difficult than addition. In contrast to Experiment 1 and in agreement with previous literature, for Experiment 2, when operation was manipulated as a within participants factor the expected effect of operation was present. This suggests that treatment type of factors routinely manipulated in arithmetic cognition research may have some effect on calibration or efficiency of arithmetic problem solution processes. Although not exactly startling in the context of our conclusions about the operand recognition paradigm discussed in Chapters 4 and 5, this may prove to be a fruitful area for future research because it seems necessary that the consequences of these effects on the currently published literature in the area should be explicitly quantified.

In conclusion, the current dissertation expanded understanding of strategy choice chiefly in four directions by 1) demonstrating that metacognitive factors affect strategy choice for basic arithmetic, 2) by demonstrating that the process of strategy selection is affected differentially by Arabic-digit and auditory-verbal input which also interacts with the operation to be solved, 3) by discovering critical limitations in a new paradigm that has begun to gather momentum as an important measure of strategy use for arithmetic, and 4) by discovering a critical failure in the sensitivity of this new method to measure the effects of factors known to influence strategy use.

Study of mathematics is classically approached from a pragmatic perspective, in the practice of math itself. Bertrand Russell (1917), a renowned early twentieth century philosopher of mathematics, famously remarked, “Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true” (p. 38, pgh. 4). Taken at face value, I believe this idea is a call to cognitive science, that it is incumbent upon us in the twenty-first century, to illuminate how we as humans relate to math ourselves. Indeed, I hope that the preceding work has made some small contribution to this important endeavour.

REFERENCES

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta, & M. Moscovitch (Eds.), *Attention and performance* (pp. 421–452). Cambridge, MA: MIT Press.

Apter, A. J., Cheng, J., Small, D., Bennet, I. M., Albert, C., Fein, D. G., George, M., & van Horne, S. (2006). Asthma numeracy skill and health literacy. *Journal of Asthma, 43,* 705–710.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130,* 544-570.

Arbuthnott, K. D. (2008). Asymmetric switch cost and backward inhibition: Carryover activation and inhibition in switching between tasks of unequal difficulty. *Canadian Journal of Experimental Psychology, 62,* 91-100.

Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition, 1*, 3-34.

Benjamin, A. S. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamneumonic judgments. *Psychonomic Bulletin & Review, 12*, 874-879.

Brysbaert, M. (2005). Number recognition in different formats. In J. I. D. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 23-42). New York, NY: Psychology Press.

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology, 54A*, 1005-1029.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53*, 1-44.

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition, 1*, 121-164.

Campbell, J. I. D. (1997). Reading-based interference in cognitive arithmetic. *Canadian Journal of Experimental Psychology, 51*, 74-81.

Campbell, J. I. D., & Alberts, N. M. (2009). Operation-specific effects of numerical surface form on arithmetic strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 999–1011.

Campbell, J. I. D., & Arbuthnott, K. D. (2010). Effects of mixing and cueing simple addition and multiplication. *European Journal of Cognitive Psychology, 22*, 422 – 442.

Campbell, J. I. D., & Austin, A. (2002). Effects of response time deadlines on adults' strategy choices for simple addition, *Memory & Cognition, 30*, 988-994.

Campbell, J. I. D., Dowd, R. R., Frick, J. M., McCallum, K. N., & Metcalfe, A. W. S. (2010). *Neighbourhood consistency and memory for number facts.* Manuscript submitted for publication.

Campbell J. I. D. & Epp L. J. (2005). Architectures for arithmetic. In J. I. D. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 347-360). New York, NY: Psychology Press.

Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition, 34*, 633-647.

Campbell, J. I. D., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition, 80*, B21-B30.

Campbell, J. I. D. & Gunter, R (2002). Calculation, culture, and the repeated operand effect. *Cognition, 86*, 71-96.

Campbell, J. I. D., Kanz, C. L., & Xue, Q. (1999). Number processing in Chinese–English bilinguals. *Mathematical Cognition, 5*, 1-39.

Campbell, J. I. D., & Metcalfe, A. W. S. (2008) Arabic digit naming speed: Task context and redundancy gain. *Cognition, 107*, 218-237.

Campbell, J. I. D., Parker, H. R., & Doetzel, N. L. (2004). Interactive effects of numerical surface form and operand parity in cognitive arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 51-64.

Campbell, J. I. D., & Penner-Wilger, M. (2006). Numeral format and response time distributions for simple addition: The μ of memory and the τ of transformation, *Memory & Cognition, 34*, 217-226.

Campbell, J. I. D., & Timm, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review, 7*, 692-699.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*, 299-315.

Carnahan, W. H. (1946). History of arithmetic. *School Science and Mathematics, 46*, 209-213.

Charette, M. F. & Meng, R. (1998). The determinants of literacy and numeracy, and the effect of literacy and numeracy on labour market outcomes. *Canadian Journal of Economics, 31*, 495-517.

Clapp, F. L. (1924). The number combinations: Their relative difficulty and the frequency of their appearance in textbooks. *University of Wisconsin Bureau of Educational Research, Bulletin No. 2*.

Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric asymmetries and task demands. *Neurocase, 1*, 121–137.

Cousineau, D., Brown, S., & Heathcote, A. (2004). QMPE: Estimating Lognormal, Wald and Weibull RT distributions with a parameter dependent lower bound. *Behaviour Research Methods, Instruments and Computers, 36*, 277-290.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33,* 219-250.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*, 83–120.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology 20*, 487–506.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2005). Three parietal circuits for number processing. In J. I. D. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 433-485). New York, NY: Psychology Press.

Finnie, R., & Meng, R. (2006). *The importance of functional literacy: Reading and math skills and labour market outcomes of high school drop-outs.* Ottawa, ON: Statistics Canada, Analytical Studies Branch.

Geary, D. C. (1996). The problem-size effect in mental addition: developmental and cross-national trends. *Mathematical Cognition, 2*, 63-93.

Geary, D. C., Widaman, K. F., Little, T. D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition, 14*, 478-487.

Geary, D. C., & Wiley, J. G. (1991). Cognitive Addition: Strategy Choice and Speed-of-Processing Differences in Young and Elderly Adults. *Psychology and Aging, 6*, 474-483.

Hecht, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, *27,* 1097-1107.

Heathcote, A., Brown, S. & Mewhort, D.J.K. (2002) Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review, 9*, 394-401.

Huizinga, M. M., Beech, B. M., Cavanaugh, K. L., Elasy, T. A., & Rothman, R. L., (2008). Low numeracy skills are associated with higher BMI. *Obesity, 16*, 1966–1968.

Imbo, I., Vandierendonck, A., & Rosseel, Y. (2007). The influence of problem features and individual differences on strategic performance in simple arithmetic. *Memory & Cognition*. *35*, 454-463.

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology, 63*, 124–138.

Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 157-175.

Koriat, A. & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 34-53.

LeFevre, J.-A., Bisanz, J., Daley, K., Buffone, L., Greenham, S., & Sadesky, G. (1996b). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General, 125*, 284-306.

LeFevre, J.-A.., Lei, Q., Smith-Chant, B. L., & Mullins, D. B. (2001). Multiplication by eye and by ear for Chinese-speaking and English-speaking adults. *Canadian Journal of Experimental Psychology, 55*, 285-295.

LeFevre, J.-A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from China and Canada. *Mathematical Cognition, 3*, 31-62.

LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996a). Selection of procedures in mental addition: reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 216-230.

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory and Cognition, 22*, 364-382.

Logan, G. D., & Klapp, S.T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 179-195.

McCloskey, M., Macaruso, P., & Whetstone, T. (1992). The functional architecture of numerical processing mechanisms: Defending the modular model. In: J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills*, Elsevier Science, North Holland.

Metcalfe, A. W. S., & Campbell, J. I. D. (2010). Switch costs and the operand recognition paradigm. *Psychological Research, 74*, 491-498.

Metcalfe, A. W. S., & Campbell, J. I. D. (2007). The role of cue familiarity in adult's strategy choices for simple addition. *European Journal of Cognitive Psychology, 19*, 356-373.

McCloskey, M., Harley, W., & Sokol, S. M. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 377-397.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134–140.

Monsell, S., Sumner, P., Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31*, 327-342.

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: Is it easier to switch to the weaker task? *Psychological Research, 63*, 250-264.

Noël, M.,-P., Fias, W., & Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition, 63*, 335-374.

Penner-Wilger, M., Leth-Steensen, C., & LeFevre, J.-A. (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory and Cognition, 30*, 1160-1167.

Pesenti, M., Thioux, M., Seron, X, & De Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience, 12*, 461–479.

Phenix, T. L., & Campbell, J. I. D. (2004). Effects of multiplication practice on product verification: Integrated structures model or retrieval-induced forgetting. *Memory & Cognition, 32*, 324-335.

Reder, L. M. & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435- 452.

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

Russell, B. A. W. (1917). *Mysticism and Logic and Other Essays*. London: George Allen & Unwin Ltd.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 3-29.

Seyler, D. J., Kirk, E. P., & Ashcraft, M. H. (2003). Elementary subtraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1339-1352.

Shrager, J. & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405-410.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General, 126,* 71-92.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modelling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Smith-Chant, B. L., & LeFevre, J.-A. (2003). Doing as they are told and telling it like it is: Self-reports in mental arithmetic. *Memory & Cognition, 31*, 516-528.

Sohn, M.-O., & Carlson, R. (1998). Procedural frameworks for simple arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1052-1067.

Stanescu-Cosson, R., Pinel, P., Van de Moortele, P.-F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain, 123*, 2240-2255.

Thevenot, C., Barrouillet, P., & Fayol, M. (2001). Algorithmic solution of arithmetic problems and operands–answer associations in LTM. *Quarterly Journal of Experimental Psychology, 54A*, 599-611.

Thevenot, C., Castel, C., Fanget, M., & Fayol M. (in press). Mental subtraction in high and lower-skilled arithmetic problem solvers: Verbal report vs. operand-recognition paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Thevenot, C., Fanget, M., & Fayol M. (2007). Retrieval or nonretrieval strategies in mental arithmetic? An operand recognition paradigm. Memory & Cognition, 35, 1344-1352.

Thioux, M., Pillon, A., Samson, D., de Partz. M.-P., Noël, M.-P., & Seron, X. (1998). The isolation of numerals at the semantic level. *Neurocase, 4*, 371-389.

Van Harskamp, N. J., & Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication. Implications for the organisation of arithmetical facts. *Cortex, 37*, 363–88.

Venkatraman, V., Ansari, D., & Chee, M. W. L. (2005). Neural correlates of symbolic and non-symbolic arithmetic. *Neuropsychologia, 43*, 744–753.

Verguts, T., & Fias, W. (2005). Interacting neighbours: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition, 33*, 1-16.

Waldrop-Valverde, D., Osborn, C. Y., Rodriguez, A., Rothman, R. L., Kumar, M., & Jones, D. L. (2009). Numeracy skills explain racial differences in HIV medication management. *AIDS and Behaviour, 14*, 799-806.

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception & Performance, 29*, 455-489

Zbrodoff, N. J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. *Memory and Cognition, 23*, 689-700.

Zbrodoff, J. N., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I.D. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 331-346). New York, NY: Psychology Press.