UC Merced

UC Merced Electronic Theses and Dissertations

Title

Math Anxiety And Metacognition in Mathematics Education

Permalink

https://escholarship.org/uc/item/26w5j72t

Author

Erickson, Shanna Lin

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, MERCED

Math Anxiety and Metacognition in Mathematics Education

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Cognitive and Information Sciences

by

Shanna Lin Erickson

Committee in charge:

Professor Evan Heit, Chair
Professor Teenie Matlock
Professor Rick Dale

2015

Copyright
Shanna Lin Erickson, 2015
All rights reserved

The Dissertation of Shanna Lin Erickson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Rick Dale

Teenie Matlock

Co-Chair *(if applicable)*

Evan Heit Chair

University of California, Merced

2015

Vita

Shanna Lin Erickson

B.S. Mathematics

Stanford University

M.S. Statistics

Stanford University

Ph.D. Cognitive and Information Sciences

University of California, Merced

(Faculty supervisor: Professor Evan Heit)

Abstract

Math Anxiety and Metacognition in Math Education
Shanna Lin Erickson
University of California, Merced
Degree: Ph.D., Cognitive and Information Sciences
Committee Chair: Professor Evan Heit

This dissertation explores the connections among metacognition, math ability, and math anxiety, predominantly within an educational context. An ultimate goal is to improve student success in math by first addressing student misconceptions of self-ability that drive study habits.

To inform later discussions of these connections, I first provide supporting groundwork research for each of these cognitive functions separately in Chapters 1-3. In Chapter 1, background information on math cognition includes discussion of basic cognitive and neurological processes necessary for numerical and mathematical abilities, numerical representations, number recognition, spatial representation, development of arithmetic skills, as well as deficits and learning disabilities in math. In Chapter 2, background information on metacognition includes a brief history of the development of metacognitive theory, philosophical debates, implications within education, and the possibility of teaching metacognition. This lays the groundwork for understanding the mechanics and influences of metacognition in general contexts as well as educational contexts specifically. Discussions of mental processes for both math cognition and metacognition in Chapters 1 and 2 allow us to gauge how metacognitive judgments of ability might be made in math specifically rather than in general contexts, as both math tasks and metacognitive tasks require overlapping resources (e.g. working memory). Chapter 3 introduces math anxiety as another possible variable that could influence both math ability and metacognitive judgments of math ability. Background information on math anxiety includes effects on math ability, cognitive ability, and causes of math anxiety. I also discuss related research in general psychological stress as well as common methodologies. This discussion of stress research introduces an alternative to traditional self-report measures of math anxiety.

In Chapter 4, I present novel empirical research that explores the connection between metacognition, math ability, and math anxiety. To extend this exploration, I also address key related concepts, including social influences on metacognition (Chapter 5) and psychological and physical stress (Chapter 6).

I conclude with a final summary as well as a discussion of future directions.

Table of Contents

Pg. 1	Chapter 1: Metacognitive Theory and Its Influences on Education
Pg. 14	Chapter 2: Math Cognition
Pg. 31	Chapter 3: Math Anxiety and Psychological Stress
Pg. 39	Chapter 4: Metacognition and Confidence: Comparing Math to Other Academic Subjects (Studies 1 and 2)
Pg. 61	Chapter 5: Metacognition, Math Anxiety, and Perceived Rank (Studies 3 & 4)
Pg. 71	Chapter 6: Effects of Psychological Stress on Metacognition (Study 5)
Pg. 78	Chapter 7: Summary, Overall Conclusions, and Future Directions
Pg. 84	References

Chapter 1: Metacognitive Theory and Its Influences on Education

Metacognition has long been an important area of research that has had growing numbers of applications within education. Metacognition plays a role in cognitive processing and is consequently intrinsic to educational pursuits, and it impacts several areas within cognitive science, such as attention, memory, perception, comprehension, and problem solving (Kitchener, 1983; Metcalfe & Shimamura, 1994). It also aids intellectual endeavors requiring complex thought processes (Schoenfeld, 1983) and also affects social behavior (Jaccard, Dodge & Guilamo-Ramos, 2005) and decision making (Cohen, Freeman, & Thompson, 1998). This paper will touch on aspects of behavioral science, philosophical methods, and computational modeling within metacognition research, and how this influences educational research.

Early Metacognition Theory: Hart and Flavell

We use metacognition not just for complex tasks, but also in daily activities, although this is not immediately obvious. Metacognition is often loosely referred to as "thinking about thinking", although the concept is far more complex than this. Josef Hart raised questions on what was to be the groundwork for metacognition in exploring the mechanisms of metamemory. An experimental psychologist, Hart questioned how information-processing systems with limited resources were able to reliably control their own memory contents. In 1965, his theory of metamemory put forth that predictive reliability of feelings of knowing in evaluating one's learning capacity in specific cases. Within a few years, John Flavell had coined the term "metacognition" in 1979. Metacognition researchers still cannot seem to agree on a common definition as they come from varying disciplines with different purposes. Because of this, aspects of metacognition are often given different terms, although they refer to the same idea, for example using the terms "self-regulation" and "executive control" interchangeably. There are usually some shared elements within definitions of metacognition, however, that borrow from Flavell's initial model. This 4-class model was comprised of metacognitive knowledge, experiences, goals (or tasks), and actions (or strategies). In this model, metacognitive knowledge refers to one's current world knowledge about the tasks, goals, actions and experiences of cognitive agents. It encompasses knowledge of beliefs about one's own nature as a cognitive processor, as well as other cognitive creatures and with their own particular cognitive tasks, goals, actions, and experiences. Metacognitive knowledge can guide cognitive processes through a deliberate and conscious memory search or through automatic cognitive processes. It can also lead to a wide variety of metacognitive experiences, described by Flavell as conscious cognitive or affective experiences related to an intellectual endeavor.

Separate from metacognitive knowledge, Flavell (1979) also discusses metacognitive awareness. While metacognitive knowledge refers to explicit knowledge about our own cognitive abilities, metacognitive awareness refers to feelings and experiences we have when engaging in cognitive processes (e.g. retrieval). Metacognitive beliefs are used to evaluate the accuracy of retrieved memories, and are also used to determine ability to retrieve a memory. Thus these metacognitive beliefs are similar to metacognitive knowledge in that they can be learned, and false memories can be corrected.

Since this first model by Flavell, most models and definitions of metacognition seem to include some version of metacognitive knowledge and metacognitive experiences, involving active monitoring and consequent regulation of cognitive processes to achieve cognitive goals. In other early models, metacognitive knowledge still generally refers to our knowledge of cognitive processes, and our knowledge about how we and others learn and process information. Metacognitive experiences are marked by use of metacognitive strategies and regulation (Brown 1987). These strategies are processes that serve to control cognitive activities, work toward a cognitive goal, and monitor the success of achieving the cognitive goal. Monitoring and regulation can take the form of checking, planning, selecting, and inferring (Brown & Campione, 1980), self-questioning and introspection (Brown, 1978), or making judgments about what one knows or does not know to accomplish a task.

Metacognitive control takes the form of decisions made based on the results of our monitoring processes. This is evidenced in what behavior or strategies one engages in as a function of monitoring. For example, if a student does not feel prepared for an exam, they may choose to continue studying. In educational settings, it is possible to measure control processes by considering allocation of study time and decisions about what content to study (Theide & Dunlosky, 1999; Son & Metcalfe, 2000).

These control processes are important for developing a theory of applied metacognition. Existence of these processes can influence human behavior and cognition, and we may further be able to improve or change control processes to aid any learning environment. Thiede & Dunlosky (1999) investigated such improvements. In their experiment, students made judgments of learning (JOLs) on paired-associate word pairs and were later given the choice to restudy work pairs. When time pressure was introduced, students chose to restudy pairs that had high associated JOLs, whereas with low time pressure, students chose to restudy harder items with low associated JOLs. This gives evidence that students made different cognitive control decisions based on context, and provides promise that we can teach students to use a flexible set of control strategies in learning scenarios. This idea of control leads to possible implications for how to apply metacognitive findings.

Conscious vs. Non-conscious Processes

Although definitions of metacognition typically incorporate metacognitive knowledge and metacognitive experiences, this leads to a certain amount of confusion once it becomes necessary to separate components of cognition and metacognition. What distinguishes a cognitive task from a metacognitive one? For example, can we term declarative knowledge as being metacognitive? Even Flavell was unable to find distinct definitions for metacognitive and cognitive knowledge, and suggested that these two might not be different. Rather, he identified use of this knowledge to separate the two. As metacognition is supposed to exercise control over thinking, cognitive strategies are simply used in reaching a goal, while metacognitive strategies are needed to monitor achievement of this goal. Thus a metacognitive experience could occur before or after a cognitive activity. Furthermore, a cognitive failure may be more likely to lead to a metacognitive process, for example, when you realize that you do not understand a concept. Metacognitive and cognitive strategies under this organization still seem as if

they could overlap unless they are distinguished by the particular use the process is used for. For example, using rereading a passage could be enacted to gain knowledge as a cognitive process, or one could reread a passage in order to gauge how much you know as a metacognitive process. Thus knowledge is metacognitive if its purpose is strategic in meeting a goal. If you are only aware of your specific shortcomings in a learning environment, this knowledge in itself is considered merely cognitive, whereas using this information to structure a study session would be metacognitive.

Repeated use of active and conscious strategies can lead these to become automatic and non-conscious. Thus conscious elements of metacognition may lessen over time, not meaning that strategies are necessarily no longer metacognitive, but are simply made more automatic and less reflective. Indeed, the need for active and conscious strategies may be lessened as the previously misguided behaviors they once mediated become more automatic (Flavell & Wellman, 1977). Even simple strategies used in approaching a difficult task may have been a result of a conscious choice, but it is possible for them to be enacted as a result of years of repeated use for other similarly difficult tasks (Nicholls & Miller, 1984). That these processes have become more automatic does not mean they no longer involved knowledge about one's own cognition. Thus is arguable that metacognitive thoughts that have since become automatic through repeated can still to be considered metacognitive, as they are still based on knowledge about one's own cognition. Automatic cognitive processes may involve knowledge and cognition about one's own cognitive behavior just as metacognitive processes do. Finding evidence of non-conscious processes presents more of a challenge, however. People are generally aware of and able to report their conscious processes used in a task. It is at least possible for people to identify the outcome of a non-conscious process, but more difficult for them to be aware of the process itself (Ericsson & Simon, 1980), making these processes nearly invisible in a lab setting even when participants are directly asked to reflect on their metacognition (Nisbett & Wilson, 1977). As this branch of metacognition has been so difficult to research, it remains unclear if thoughts that were once metacognitive that have become unconscious and automatic should still be considered metacognitive. Some distinction between these automated processes and other more obvious forms of metacognition should be made, perhaps based on the degree to which metacognitive processes involve an awareness of oneself as an actor in their environment, as they actively store and retrieve information. But until there is a better solution to studying non-conscious thought than unreliable self-reports, it seems reasonable and practical to consider only active and conscious thoughts as a metacognitive (e.g., Paris & Winograd, 1990; Borkowski & Muthukrishna, 1992; Carr, Alexander, & Folds-Bennett, 1994). These are likely able to be mediated by the person experiencing them, and are potentially more accurately reported, making them easier for researchers to study.

Developments Since Flavell

Kluwe (1982) added to Flavell's model by identifying two distinct components of metacognitive activities. This first is having knowledge about one's own thinking and the thinking of others, Kluwe linked to declarative knowledge in long-term memory. The second component is monitoring and regulating and acting as an agent in the course of

one's own thinking. This component was considered procedural knowledge stored in the processes of a system. The significance of this distinction was used to determine what is actually metacognitive. Both knowledge in long-term memory and stored processes of a system can be found at metacognitive and cognitive levels. Kluwe argued that at cognitive levels, stored data may consist of domain knowledge only (e.g. knowledge about social interactions or knowledge about science), whereas stored processes may consist of solution processes for specific problems. Metacognitive procedural knowledge he further termed as processes that monitor the selection and use as well as the effects of solution processes (Kluwe, 1982). Kluwe also called this procedural knowledge "executive processes", which involve both monitoring and regulating other thought processes. This corresponds with Flavell's (1979) metacognitive strategies and Brown's (1978) metacognitive skills. Executive monitoring processes involve one's decisions that help to identify the task on which one is currently working, as well as evaluates progress and predicts the outcome of that progress. Executive regulation processes further involve one's decisions that aid in allocating resources to a current task, determine the order or actions needed, and determine the pace needed to be effective. While creating a general distinction between procedural and declarative knowledge Kluwe also helped to emphasize the importance of metacognitive research as a way to gain greater understanding of people as self-regulatory, and capable of assessing themselves and others in enacting behavior toward specific goals.

This emphasis is critical. Our thinking is not something that happens independently from us or our own guidance. Rather, it is can be controlled deliberately (Kluwe, 1982). People who realize they are agents of their own thoughts are more likely to overcome shortcomings of skill, as they will not be convinced that they will forever lack an ability with no possible way of improving this ability. Those who are instead convinced that there is nothing they can do to improve their current knowledge or ability will have little motivation to attempt a solution, and even less likely to regulate this process.

Paris and Winograd (1990) similarly introduced self-efficacy as part of metacognition, and believed that metacognition incorporates self-appraisal and self-management of cognition. Self-appraisals are reflections about knowledge states and abilities, and affective states about knowledge, abilities, motivation, and learner characteristics. Self-management refers to processes that regulate aspects of problem solving in tandem. These two components cast learners as being required to be actively involved in their thinking.

These definitions provide us with an idea of the basic processes that are typically thought to comprise metacognition. Flavell's first model has been more precisely considered and defined, although not all metacognition researchers yet agree on all less-studied aspects of metacognition. At the very least, there does seem to be general inclusion of knowledge of one's knowledge, processes, and cognitive and affective states, and the ability to consciously and deliberately monitor and regulate one's knowledge, processes, and cognitive and affective states.

Philosophical Debates

Considering metacognitive research as a whole, we have not been able to arrive at a single encompassing definition. This is not surprising, considering the variety of methods, background assumptions, and experiments covered by many subfields within cognitive science. Some give focus to underlying mechanisms while others consider only philosophy of mind and metal states that regulate and rationalize human behavior. While disparate, functionalist theorists and philosophers of mind support ideas that are not necessarily contradictive. Rather, they help to form a more complete view of metacognition from the point of cognitive science.

Ideas put forth by Hart and Flavell are still widely regarded and have provided much evidence for the mechanisms guiding memory retrieval, reasoning, and learning. However, current theoretical debates surround the two camps of thought. In particular, do the two alternate meanings of metacognition use the same process or processes? In Flavell's understanding, metacognition is thought about thought, and as such attributes thoughts to one's own thought, guided by beliefs about universal properties. However, these properties are the same for self as it is for others, which could point to a lack of theoretical distinction between knowledge about self, and knowledge about others. Examples of this can be found in in the work of developmental psychologist such as Alison Gopnik, Peter Carruthers, and Josef Perner, who use the term metacognition to refer to forms of thought attribution, or social cognition, that they claim to have to engage a theory of mind ability.

In contrast, Hart's approach to metacognition does not address current or past attitudes about a self-belief or desire. Rather, it evaluates present cognitive dispositions or outputs of self, then form epistemic and conative commitments. This evaluation does not have to incorporate beliefs or metacognitive knowledge about how memory should be controlled. It simply uses cues such as feelings of knowing or feelings of ability that will control one's activity. Thus most researchers who support Hart's theories consider cues to be related to feelings, rather than conceptual content about what constitutes memory. This leads to the question of whether cues are only useful if the person is able to interpret them in terms of mental content. Another question is whether we can speak of epistemic commitment considering the lack of formed judgments concerning the correctness of suppositions retrieved from memory. Addressing such questions further supposes the operation under a single meaning of metacognition.

In the theoretical philosophy of mind debate surrounding these two camps, one can criticize Hart's theory that one needs to form beliefs about one's propositional attitudes with their associated contents. This is necessary in order to monitor one's mental dispositions or support one's attitudes. Under this assumption, metacognition as monitoring cognition is a simple consequence of metacognition as knowing. Thus metacognition could be inclusively defined as the capacity to attribute mental states to self or to others, making evaluative states a particular case of this general ability. In contrast, one can criticize that it is conceivable that metacognition is related to one's own abilities and their evaluation, and that at least part of this evaluation is carried out without even a theoretical knowledge of mind. If metacognition turns out to be the ability for self-evaluation based on procedural knowledge of knowing how rather than knowing about, then it one could define metacognition as referring only to the capacity of self-

evaluating one's own thinking. This could in turn be rejected by opposing theorists as such a definition would prevent them from using attributions of others in evaluating metacognitive abilities.

Rather than choosing one of these opposing theories to support in this paper, I will rather consider research from both points of view in an effort to encompass more of the work that has contributed to what we know about metacognition. Joelle Proust presents a provisional definition that is neutral between both the inclusive and exclusive camps: "Metacognition is the set of capacities through which an operating cognitive subsystem is evaluated or represented by another subsystem in a context-sensitive way." Admittedly, this definition falls short for both theories, but it can serve as a way to consider both areas of research equally within this paper. From this neutral understanding of metacognition, we may further discuss past and growing areas of research.

Metacognitive Research

Early descriptive studies of metacognition have since evolved to include empirical measures, using a growing number of methodologies (e.g. Kluwe, 1982; Schoenfeld, 1987). Three categories of research have arisen from the days of Flavell. The first is cognitive monitoring, in which studies examine people's knowledge of their knowledge and thought processes and their accuracy in monitoring the current state of these (Kluwe, 1982; Schoenfeld, 1987). Examples of studies include ones that examine predictions of knowledge stored in memory, and allocation of study based on judgments of knowledge that is not stored in memory (Schneider, 1985). This often involves self-reports during a task (Cavanaugh & Perlmutter, 1982). The second category of metacognition research examines regulation of one's own processes towards adapting to changing necessities in completing a task. Studies of this kind often incorporate a training task followed by strategy transfer task (Schneider, 1985). This more dynamic type of study requires participants to master a strategy then decide whether or not this strategy is applicable to a subsequent task difference from the first. They may use the original strategy, modify it, or even choose yet another strategy that they deem more appropriate for the transfer task. The third category of metacognition research includes studies that examine regulation and monitoring, in which participants monitor available information during thinking then use this information to regulate memory processes (e.g. Kluwe, 1982; Schoenfeld, 1987; Schneider, 1985). These studies are used to consider memory strategies and how strategies can be used to improve performance (Schneider, 1985), and are used to reveal what people know about memory relevant to performance (Cavanaugh & Perlmutter, 1982).

Educational Research

Explanations of the phenomena within memory and reasoning, theory of mind, epistemological understanding, knowledge acquisition and problem solving all require an understanding of metacognition. This high level of influence on so varied an array of research areas can be encompassed by considering metacognition to progress over a developmental timeline. Kuhn (2000) proposes that these early foundations begin when children first become aware of mental functions, such as how one comes to know

something, and progresses as their development starts to include increasingly complex metacognitive capabilities. These capabilities become more explicit in nature over time and are enacted under conscious control. The eventual understanding that knowledge results from human knowing is the basis for epistemological and metacognitive thinking (Hofer & Pintrich, 1997). These early foundations also influence scientific thinking (Kuhn & Pearsall, 2000), which is critical to many aspects of educational learning. Scientific thinking takes into account new evidence that leads to new knowledge. This requires consciously deliberate, rule-governed, and hence, metacognitively controlled processes.

Two components of metacognition are of particular importance in education: the ability to monitor what you know, which acts as a basis for predicting retention, and the control processes that are used to enact study activities (Nelson & Dunlosky, 1991). Consideration of these two components constitutes much of behavioral research within education in how students regulate and monitor their learning and in how they enact learning strategies based on their self-assessments. Students need to use metacognitive control in gauging what they know and deciding what study methods to use (Thiede, Anderson, & Therriault, 2003; Metcalfe, 2009). This process is constantly changing, as students adapt their behaviors in monitoring a learning goal. Self-regulation is necessary for this process (Kornell & Bjork, 2007), thus students must select from a variety of strategies, enacting these strategies in goal-directed activities, and monitoring their progress in using these strategies.

Within education, one can also further consider two types of metacognitive abilities, one for evaluating procedural knowledge and one for declarative knowledge, which Kuhn (2000) has termed as metastrategic knowing and metacognitive knowing, respectively. Considering both these functions helps explain when cognitive development occurs and when it does not.

Success of metacognition affects students' academic performance (Hattie, Biggs, & Purdie, 1996; Paris & Paris, 2001; Coutinho, 2008), as well as their ability to communicate what they know about a particular problem. Being able to communicate their level of understanding to instructors is crucial to the learning process. It guides how classroom and self-study materials are constructed, and can affect what strategies students are taught for study and examination. Metacognition has been shown to play a crucial role in gauging one's own knowledge (Sperling, Howard, Stanley & DuBois, 2004), including specific academic domains such as reading comprehension (Pressley, 2002), math (Pugalee, 2001), science (Schraw, Crippen, & Hartley, 2006), and writing (Pugalee, 2001).

Many researchers who advocate the relevance of metacognition in educational settings have since moved to conduct more practical rather than theoretical studies, and explore how metacognitive processes can facilitate learning. Of particular importance in education is self-regulation in one's ability to monitor knowledge and processes. In this educators are united, considering self-regulating learners to be the most effective learners (Butler & Winne, 1995). Instrumental to self-regulation is being able to accurately self-assess one's own knowledge (Schoenfeld, 1987). Students cannot self-direct their learning if they have no starting point. They must know the current state of their own knowledge. Thus the instruction of self-regulation for complex tasks is of great interest

to educators. It has been shown that students in elementary grades can monitor current knowledge and processes in memory, and this ability improved with age (Bisanz, Vesonder & Voss, 1978). When judging memory monitoring ability, one must take into consideration more than just age. Nelson and Dunlosky (1991) used judgments of learning (JOLs) in gauging accuracy of memory monitoring for recently learned knowledge. They found that JOLs increased in accuracy after a delay rather than immediately after learning. This highlights the importance of considering whether it is long- or short-term memory that is being monitored in metacognition research.

A study by Lodico, Ghatala, Levin, Pressley, and Bell (1983) was used to determine if instruction of general principles of strategy monitoring would inform student's regulation of strategies on other related tasks. They thought it would be necessary for children to be made aware of the relationship of a specific strategy in improving performance if they were to continue using the strategy on their own on related tasks. Results were that a greater proportion of second grade students the children who were taught to monitor strategy effectiveness than of those who were not so trained could identify that a specific strategy improved their performance. They were also better able to choose a more effect strategy amongst differing choices, again identifying their improved performance as a reason for their choice. Similarly, other studies have shown that students can be taught how to monitor strategic behavior in improving their performance, thus improving their regulation of strategies. This gives promise for the possibility of teaching metacognitive awareness.

Other studies have examined how students regulate their strategy selection using information they monitor while strategies are under way. One such study is one by Pressley, Levin, and Ghatala (1984). Students were taught to use sentence-elaboration and repetition strategies to learn vocabulary words, and were later asked to choose the more effective strategy when given a list of new vocabulary words. Students were either allowed or not allowed to practice the presented strategies before being asked to make a strategy choice. All students had to make a choice for strategy, however. Results showed that those without practice were more likely to choose a strategy recommended by the experimenter. Those who had the opportunity to practice strategies were likely to choose a more effective strategy regardless of a recommendation given by an experimenter. This shows students' ability to gain awareness of the effectiveness of strategies and their relative usefulness in particular scenarios.

Past findings indicate that females generally lag behind their male counterparts on standardized test performance in math (Brown & Josephs, 1999). This is particularly true among high school and college students (Hyde & Linn, 2006). This gap does appear to have narrowed in recent years (Else-Quest, Hyde, Shibley, Marcia, 2010). However, attitudes toward math between genders still follow differing patterns, and females are more likely to feel intimidated by math than are males (Jakobsson, 2012; Brown & Josephs, 1999). This lack of confidence often leads to a self-fulfilling lag in performance (Brown & Josephs, 1999; Kiefer & Sekaquaptewa, 2007) that can lead to gaps in performance between genders. These gender differences in academic performance might be due to differences in metacognition. In a study by Bidjerano (2005), gender differences in the use of self-regulated learning strategies were examined. Results showed that females were more effective in using metacognition, and strategy choice differed

between genders. Metacognitive strategy choice likely contributes to differences in performance, and findings of this study are consistent with findings from previous research that female students exhibit better strategy use than male students (e.g. Zimmermann & Martinez-Pons; 1990; Wolters, 1999). Bidjerano also posit that female students might have been more reflective of learning experiences and thus more aware of the strategies they use in learning, as well as a greater willingness to report of the use of these strategies. These differences in self-regulatory skills might further be highlighted by stereotypical beliefs about genders that students hold (Pajares & Valiante, 2002).

Teaching Metacognition

Metacognition is now commonly discussed in educational settings and how it can be used in student writing, problem solving, and learning across different domains (e.g. writing vs. mathematics). For instance within writing, strategy instruction can improve quality of writing, focusing on student monitoring of performance during a writing task. The goal of this is self-regulated strategy development and importance of metacognition in understanding writing skills. In science and mathematics, students rather will need metacognitive skills that will aid their regulation of scientific inquiry.

There is no lack of work that emphasizes the importance of student metacognition. Flavell (1979) in his early work already addressed the possibility that children, and even adults, could be taught how to develop their metacognitive abilities to better aid them in formal educational settings. However, there is no corresponding literature for how teachers can incorporate metacognition into school curriculum. In order to teach metacognition, teachers must themselves be aware of metacognitive strategies, something that is not included in teacher training programs (Veenman et al., 2006). Teachers must provide strategic instruction and have the ability to be flexible in their feedback to students. This demanding process requires teachers to be aware of the strategies they are trying to convey and also instructional strategies they themselves can use, and current teacher training programs do not prepare teachers for this type of teaching. In a study by Duffy and Roehler (1987), the effect of teacher training in providing explicit reasoning explanations and mental processes in reading comprehension was examined. This approach did not teach individual strategies, instead it focused on helping students to frame reading as a problem solving task that required the use of strategic thinking. Additionally, students were encouraged to think strategically about solving reading comprehension problems. Implementation of this approach required twelve hours of directed teacher training on how to teach within this framework, and how to provide explicit explanations, to analyze skills needed, and to frame the strategies as problem-solving. These trained teachers were later compared with those who had not received this specialized training, over the course of the year-long treatment. The students of trained teachers showed higher levels of awareness of specific reading strategies and a greater awareness of the need for strategies when reading when compared to those students taught by teachers that did not have the strategy-teaching training. There was no significant difference between the two conditions in ability to use base reading skills. However, the students in the treatment group were better able to use strategic reasoning and had greater awareness of the need for this strategic reasoning

when reading in comparison to the control group. These results indicate that training teachers to incorporate reading strategies into instruction can aid student metacognition.

Borkowski & Muthukrishna (1992) have since argued that metacognitive theory has great potential for improving classroom environments by encouraging focus on strategic learning. Paris & Winograd (1990) further believe that students can improve their learning by being aware of their own thinking during tasks such as reading, writing, and problem solving, and that teachers can promote this awareness using direct identification of effective problem-solving strategies and discussions with students about cognitive characteristics of thinking. Brown (1997) has suggested that metacognitive development can begin at an external, social level in what he terms a "community of learners" in a curriculum that emphasizes reflective activities that will eventually become self-reflective practices.

It is of particular importance to devise ways of incorporating metacognition into classrooms. It has been shown that students are capable of improving their metacognition in learning scenarios, but there is no structured instruction of metacognitive techniques in schools. Implementation of such a structure in schools holds much promise in improving the learning experiences of our students. Any improvements in metacognition would allow students to better judge what they know and how well they will be able to recall information.

Computational Modeling of Metacognitive Ability

More recent metacognition research in educational domains explores computational models of concepts related to metacognition. However there does not yet exist a model of metacognition itself. There do exist models of specific elements of metacognition, however. In time these might be synthesized into a more complete model of metacognition.

Son & Sethi (2006, 2009) present a model of study time optimization, in which time allocation strategies are dependent on characteristics of the learning environment (e.g. time pressure). Students often spend most time on tasks thought to be most difficult or farthest from a learned state (Dunlosky & Hertzog, 1998). However, time is also often allocated to tasks of intermediate difficulty, while challenging tasks will only receive time when time pressure is not binding (Metcalfe, 2002). From these earlier studies, the idea of optimality has been mentioned, but not directly addressed. Son & Sethi's model furthers this area in addressing students' ability to use good metacognitive control in a practical setting. In allocating study time for various educational tasks, students must exercise monitoring and control in deciding how to enact study strategies, such as how to space their study time, when to self-test. From their model, it is clear that students allocate study time at least partially based on their feelings of knowing and their self-evaluation of ability. Once they feel they have mastered a topic, they move on to another. Should students possess inaccurate self-evaluations of ability, this will lead to non-optimal allocation of study time and consequently inadequate preparation for class assessments (cf. Townsend & Heit, 2010; 2011). One shortcoming of their model, that they acknowledge, is the lack of consideration memory decline over time. A way to improve Son & Sethi's initial model would be to incorporate memory models that take into account an attrition of knowledge over time (forgetting curves). For example,

switching between tasks and returning to an original task is shown to be beneficial in their model of time allocation, but under what time restraints does this hold true?

Others have addressed the effects of tests on learning and forgetting. In two specific studies, restudying was as beneficial or more beneficial than testing when retention was measured after several minutes, but testing was more beneficial than restudying after a period of two days (Wenger, Thompson, & Bartling, 1980) or 7 days (Wheeler, Ewers, & Buonanno, 2003). From this, Wheeler et al. (2003) proposed that testing reduces the rate of forgetting over a matter of days but does not increase the original extent of learning. Carpenter, Pashler, Wixted, & Vul (2008) further developed three experiments in which they investigated whether memory tests improved learning and reduced forgetting more than additional study does. After learning obscure facts through either the completion of a test with feedback or receiving an additional opportunity for study. Recall was then tested after 5 minutes, or 1, 2, 7, 14, or 42 days. An ANOVA showed that testing, rather than restudying, enhanced overall recall more. Furthermore, a power function showed that testing also reduced forgetting more than did restudying. Their results are helpful in an educational context in showing that tested information will be better retained over time, more so than information that has been restudied. Taking this one step further, students who are aware of this can enact self-testing strategies that will be beneficial for long-term information retention, though this requires students to independently realize when it is appropriate in their study progress to self-test.

Markant and Gureckis (2010) argue that self-directed learners can choose information relevant in testing a hypothesis. They developed a computational model that tracks the advantages of self-directed learning, for which learners can control flow and access of information, over "yoked" learning, for which learners have no control over information access and rather receive information passively from experimenters. Simulations reproduced this difference in performance between these two styles of learning. Simulations further showed that for more passive learning, there existed a bias toward considering rules along a single dimension. Furthermore, self-directed participants were more likely to consider a greater number of hypotheses about the category structure that they were investigating. Self-directed behavior can in turn be affected by individual rule-preference, however, especially when target rules do not match expectations of the learner (Druck, Settles, & McCallum, 2009). Markant and Gureckis conclude that self-directed selection of information is not a subset of passive learning that simply requires an initial action of the learner. However, they posit that there are implications for what is learned and what can be learned. Self-directed learning seems to improve rate of learning and memory retention, and can lead to increased development of patterns of inference and generalization and resulting decision-making strategies.

Recent developments in machine learning mirror these findings surrounding self-directed learning. These learning models, rather than utilizing passively fed training data, incorporate algorithms that must instead gather their own training data (Druck et al., 2009). The goal has been to develop machine learners that require less training than passive ones, yet with equal performance to an acceptable level of error. And such models have repeatedly shown that this is possible (e.g. Dasgupta, 2010; Tong & Koller,

2001). In general, more data helps learning models identify the threshold of error, for example, in categorizing items into two types. But many observations will not provide any additional information. More efficient models are capable of querying specific items, which effectively renders the threshold estimation problem into a binary search. Thus, an ideal learner would be faster if they are able to ask specific and targeted questions rather than through random sampling. This has further been shown for human learners (Castro et al., 2008; Markant & Gureckis, 2010), and this basic idea has been applied to more complex learning tasks (Druck et al., 2009).

These models take for granted students' accuracy of their own self-evaluations. But people's evaluations of their self-ability are often inaccurate, sometimes egregiously so (Kruger & Dunning, 1999). People have shown themselves to be quite unskilled and unaware of this fact in predicting performance on laboratory tests ranging from logical reasoning to grammatical knowledge and sense of humor. Studies with academic content in classroom settings also exhibit this phenomenon (Miller & Geraci, 2011; Maki & Berry, 1984). This phenomena has also been documented through the use of regression models (Erickson & Heit, 2013; Krueger & Meuller, 2002). Miller and Geraci further demonstrated that the lowest performers were overconfident in exam score predictions, but they were also less confident in these predictions than were the highest performers. Thus, although the unskilled might be more aware than once thought, they still demonstrate overconfidence nonetheless. Furthermore, while people with high performance might demonstrate slight underconfidence, people with lower performance have even more exaggerated overconfidence. Still others (Burson, Larrick, & Klayman, 2005) have found that both the skilled and unskilled can be unaware of their actual ability. They further investigated the theory by exploring tasks of differing difficulty. Through a Monte Carlo simulation of 100 participants, they propose a model that depends on raw ability, bias derived from the difficulty of the task, and error from a normal distribution whose variance demonstrates noise in participants' estimates of their ability. They show that tasks of moderate difficulty produce low overall bias or difference in predicted accuracy between the skilled and unskilled. More difficult tasks, however, produced a negative bias that shows low performers to be more accurate in their judgments than are high performers.

None of these current models yet provide a direct model of metacognition, but rather provide of separate components that contribute to metacognition within education. However, these separate models hold promise for the development of more comprehensive models of metacognition.

Conclusion

Metacognition remains one of the most promising areas of research within the educational context. Applications have included problem solving, reading, writing, self-regulation, technology, comprehension, metacognitive strategies, self-regulated learning, tutoring, measurement, as well as its use within individual academic domains such are writing, reading, science, and mathematics. A metacognitive approach to education makes an easy fit with the number of fundamental components of metacognition that are readily applicable to student tasks (e.g. monitoring and regulating cognition). Through further study we can hope to further how we conceptualize metacognition within

education, how we can assess it, and how we can develop it for both teachers and students.

Chapter 2: Mathematics Cognition

This chapter explores foundational cognitive and neurological processes that govern numerical and mathematical abilities. Several components affect the learning of mathematical concepts and procedures, including cognitive development, neurological development, linguistics, cognitive and educational psychology. Mathematical cognition can be broken into smaller components of varying complexity, from how the mind represents a single number to calculations requiring greater mathematical ability. I will consider several of these components, including numerical representations, number recognition, spatial representation, and development of arithmetic skills, and include a discussion for deficits in these components that arise from mathematical learning disabilities. I will address some language structure attributes of mathematical cognition, particularly within number recognition. In addition, I will discuss neuroimaging evidence for numerical processing in the parietal lobe. For these chosen areas of discussion, I primarily consider early learners of math, whose success in formal mathematics education is based on first developments of mathematical cognition but will also consider adult development for the sake of contrast in exploring the reasons behind early development. I also consider cognitive engineering applications for later learners who have entered formal schooling. Through discussion of these topics, I plan to explore the diversity of research surrounding mathematical cognition and through this diversity to highlight the importance of understand mathematical cognition to support developments in improving mathematics learning and education.

Numerical Representation and Recognition

Cognitive number processing has only recently been studied in an effort to identify distinct numerical abilities and what representations are involved for differing numerical abilities (Butterworth, 1999; Dehaene, 1992). Supplementing this work were explorations into developmental disorders surrounding these same tasks. Still more recently, Fayol and Seron (2005) have tried to gain insight for mental representations that allow for elementary processing of numbers. In particular, they seek to distinguish between numerosity, numerical notations (or symbolic codes), and internal or mental numerical representations (also referred to as semantic representation of numbers). They define numerosity as the numeric properties of a set of items in the real world. Numerical notations instead are used in order to represent numerical information external to a person (e.g. Roman numerals). Finally, internal or mental numerical representations are used for entities that are internal to a person, and this can refer to either systems of numerical notation or to numerosity of sets. These sets can be of real objects or mental events.

So what constitutes numerical representations? We know that infants become aware of quantity very early in development. They can distinguish between groups of objects or counters if the quantities are small (Antell & Keating, 1983; Starkey & Cooper, 1980; Strauss & Curtis, 1981). We also see this in their ability to discriminate and match numbers of events (Canfield & Smith, 1996; Sharon & Wynn, 1998) based on quantity. But are these representations strictly numerical? This would imply that these discrete representations are innate and solely used for number processing (Wynn, 1998). However, there might instead be a system for continuous quantities that only later

develops the structure for discrete numerical processing. Simon (1997) further suggests that the system might act for both number and quantity.

Much research, through habituation tasks, has found that preverbal infants can possess a mental representation of small quantities. Due to the high level of error in children's counting performance (Gelman & Gallistel, 1978), it was first assumed that knowledge of numerosities stemmed from formal education. Further research that such error only occurred for the case for quantities higher than three, and that children were able to distinguish between two numerosities when the difference between the two was sufficiently large (Antell & Keating, 1983).

In general, these experiments combine number of items with related continuous measures such as surface and volume (Stakey & Cooper, 1980). Feigenson, Carey, and Spelke (2002) further controlled both numerical and continuous dimensions, and they showed that children do not specifically respond to numerosity in either habituation or transformation tasks. From this they hypothesized that infants observed a limited number of objects, they create a mental model for each of the objects. They suggest that the model stores concrete information about each object, such as color and shape. This would allow for numerosities of two or three objects, as well as larger quantities such as 5 or 6 that would otherwise not be possible under their parallel processing capabilities. Feigenson, Carey & Spelke (2002) further showed that infants 10-12 months old also grasped the concept of more than or less than. These experiments show that young infants are able to utilize two different systems for processing quantities, one for absolute limited-size sets, and another approximate evaluation of large quantities along a continuous dimension.

Does this evidence suggest that symbolic arithmetic emerges? Carey (2001) suggests otherwise, suggesting that language acquisition and symbolic notation result in new representations, bringing into question the relations between these early and subsequent representations. Whalen, Gallistel and Gelman (1999) sought to find evidence of these systems by contrast in adults, asking subjects to quantify tapping movements or sets of tones. These tasks were performed under time constraints so that verbal counting strategies were not possible. Their results show that adults possess a representation of quantity that acts the same as that of an infant, both for estimation nor production. Barth, Kanwisher, and Spelke (2003) instead suggest that this representation is abstract, and that normal subjects were capable of approximate comparisons between sets of numerosities. This held true of both visual and auditory tasks. Subjects in these studies similarly constrained response time so that symbolic notation was not possible.

These results indicate that when adults process numerosities in situations that prevent symbolic representation, they use an analogue representation of numerosity. This ability extends to comparison of numerosities across modalities. Attributes of modality, such as duration, rate, or area, may instead provide people with cues in creating abstract representation. Thus number semantics are provided by processes that associate numerosity and symbolic notations with internal representations. However, it is still unclear how preverbal representations of numerosities are connected to symbolic representations. It appears that children possess two different systems for representing numerosities, but these systems have limitations. Limited-size numerosity can only be used for small sets and is not capable of representing cardinal values. Accuracy of large-

size numerosity for continuous dimensions is eventually rendered inaccurate by an upper limit of error that increases with larger and larger values. The transition from protonumeric representation to symbolic representation does not occur rapidly. The lag in time between when children can learn how to recite verbal number sequences and when they start to use this skill for cardinality judgments and calculation is quite long. It seems unlikely that the protonumeric system implicitly contains all the components of symbolic number and arithmetic processing, or that there is a partial discontinuity between these systems.

Thus far, numerosity has taken for granted that numbers can be presented in different formats, such as through numbers of taps or through a series of tones. Numbers can still further be represented in still other formats, including numerals, sequences of letters or words, and a myriad of other ways. How then are these different formats processed?

Numbers very simply serve to represent numerosities or quantities. We have seen that adults can differentiate between small and large numerosities. Time given for perceiving these numerosities plays a crucial role in this ability. Ability to determine both numerosity and difference in numerosity is greatly inhibited by shortening display time. For instance, people are much more able to accurately determine if they were shown four dots on a screen when given 50 milliseconds than they are at determining if they were shown ten dots (Kaufman, Lord, Reese, & Volkman, 1949; Jensen, Reese & Reese, 1950). More time is needed for actually counting individual items. Determination of large numerosities is further aided by visual patterns. For instance, it is easier to ten dots arranged in two rows of five rather than ten dots randomly scattered dots (Mandler & Shebo, 1982; Wender & Rothkegel, 2000).

Other repeated behaviors also arise when people are given limited time that prevents counting or use of other symbolic representations. When given more than four items at a time, participants often underestimate number of displayed elements, and this error increases with numerosity (Krueger, 1982). The larger the numerosity displayed, the greater the underestimation error. Another finding is that adults tend to estimate numerosities with greater variability when using a basis of analogue displays (van Oeffelen and Vos, 1982). Van Oeffelen & Vos showed participants displays with random configurations of dots, and asked participants if there were 12 elements on the display. Participants in this study reported an unusually high rate of affirmative responses when different numbers of elements had been displayed. An experiment by Whalen, Gallistel, and Gelman (1999) further shows that variability in estimates in general increases with larger target numbers.

Clearly mathematically literate adults can perceive numerosities smaller than four in an analog format. For larger numerosities, they must start to count or they instead make estimates. Because these two behaviors also characterize animal cognition, some theorize that they are an innate and preverbal numerical system. Still others explain empirical evidence by perceptual factors unrelated to numerical cognition. That people spontaneously and consistently start to count numerosities larger than four provides evidence for the importance of symbolic representations for numerical cognition. How then do we process these symbolic representations? Research on the processing of written verbal numerals seems to align well with what is known about the processing of

visually presented words (Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990). There is further evidence that the meaning of verbal numerals is not activated fast enough to influence response for tasks such as number naming and size of number words. This in turn supports the assumption of nonsemantic models of word processing in both localist and distributed models. A large majority of existing models for recognition of number words contain an orthographic lexicon with localist representations. Others (e.g. Campbell, 1994) carry a multiple encoding view, in which numbers are simultaneously encoded in multiple ways. In this model, number recognition depends on the pattern of coactivation of the different codes instead of activation of a single localist code. Seidenberg and McClelland's (1989) contrasting distributed model of visual word recognition instead assumes that word knowledge does not begin when activation of an entry in the orthographic lexicon exceeds a threshold. Rather, it comes about from the co-activation of processing units that encode the phonological, orthographic, and semantic properties of words (Van Orden, Pennington, & Stone 1990). A visual word activates a number of orthographic units representing the sequence of input letters, and this activation then spreads to the semantic and phonological units that are connected to the activated orthographic units. This then feeds back until a stable state is reached. Various units are also no longer attached to single words. Rather, each unit is activated by many words, and the identity of a word is determined by a pattern of activations across multiple units.

Most models for the recognition of number words contain an orthographic lexicon with localist representation. However Campbell (1994) notably deviated from this framework, instead embraced a multiple encoding view. In this model, numbers are simultaneously encoded in several ways via activation. For this, number recognition depends on pattern of coactivation of the different codes instead of on the activation of a single localist code.

A further investigation within visual word recognition is for the level of importance the meaning system within the language architecture. Interestingly, researchers have found that meaning can be ignored in their explanations of results. This has been shown for the naming of visually presented isolated words (word naming) and also for deciding if a string of letters forms a correct English word (lexical decision). Word naming tasks are typically thought to have three routes. The first is a direct conversion from letters to sounds. This allows us to name unknown sequences of letters (e.g. nonwords). The second route takes an orthographic input lexicon to a phonological output lexicon. This allows us to correctly pronounce irregular words (e.g. dough). The third route similarly starts with an orthographic input lexicon, but then proceeds through the semantic system to the phonological output lexicon. This last route is assumed to be slow and therefore does not affect performance. Furthermore, lexical decision times have been explained by focusing on a task within the word-form lexicon, with little contribution from word meanings.

Verbal numerals typically assume asemantic routes in visual word processing. For example, Fias, Reynvoet, and Brysbaert (2001), designed a task that presented participants both a verbal numeral and an Arabic numeral simultaneously. Participants were then asked to name the verbal numeral and ignore the Arabic numeral. Participants showed a high success rate for this task. The naming latencies were the same when the

Arabic numerals referred to different magnitudes than the verbal numerals (e.g. four, 8) just as when they referred to the same magnitudes (e.g. four, 4). However, when participants were asked to indicate whether the verbal numeral was odd or even, a task that requires participants to consider the meaning of the verbal numerals, they were faster in their responses when both numbers referred to the same magnitude than when they had to compare differing magnitudes.

Reynvoet, Brysbaert, and Fias (2002), by contrast, also show that the route that utilizes the semantic system is slightly slower in the naming of words. Thus meaning is still important in the three-route model. They showed that the naming of verbal numerals was primed by Arabic when values were close. For example, participants named the target word "five" faster when primed with 4 or 6 than when primed with 2 or 8. A further experiment showed the same effect with masked primes. This cross-notation priming effect indicates that the number magnitude route can easily be preactivated in order to find semantically mediated effects in naming verbal numerals.

Cappelletti, Butterworth, and Kopelman (2001) found other evidence for the significance of the semantically mediated route in naming verbal numerals. They document a case of a semantic dementia patient who had a vastly lowered word reading ability. This patient would read only 21% of words with regular letter-sound mappings and only 12% of words with irregular mappings. However, the same patient could read verbal numerals without difficulty. This spared numerical knowledge is also common in Alzheimer's disease and supports the finding that numerical knowledge is represented separately from other types of semantic knowledge (e.g. Pesenti et al., 2000).

Another related discussion of language structure surrounding the visual word recognition literature is how morphologically complex words are recognized. For number reading, only verbal numerals from zero to twelve are clearly monomorphemic, consisting of only one meaning unit. Relatedly, words such as "thirty-six" are clearly polymorphemic. However, words such as "thirteen" or "fifty" are ambiguous. Since these differ from the original words (i.e. not "threeten" nor "fivety"), if it not clear whether they can be considered polymorphemic. There are types of polymorphemic number words that are more consistent. Compound words that combine two or more words (e.g. "thirty-six") are one such type. Another type are those created by adding a suffix to a simple number word (e.g. "sixty"). Morphologically complex words would potentially be processed in two ways (Bertram & Hyona, 2003). One possibility is that they can be stored as a whole in the mental lexicon. Alternatively, they can be decomposed into their constituents which then are used to determine meaning. There have been varying theories of the relative importance of these two ways and what factors should go into determining this. Some proposed variables have been explored are semantic transparency, length of the constituting words, and word frequency (Brysbaert, 2005). Brysbaert suggests that words that are morphologically complex are more likely to be stored and retrieved as a whole when the semantic relation between the word and constituents is not clear, when the complex word occurrence is frequent, and when the complex word is short. From this, Brysbaert proposes that these factors allow for prediction that verbal numerals that are high-frequency and short, without clear relationship between constituents and the complex word are more likely to be recognized as a whole. By contrast, verbal numerals that are lower-frequency and semantically

transparent are more likely to be decomposed into constituents. Similarly verbal numerals that are low-frequency and long are more likely to be decomposed into constituents to be processed.

Thus there is evidence that the meaning of verbal numerals is not activated fast enough to influence response for many tasks (e.g. determining size of number words, and number naming). This supports the assumption of both localist and distributed models that nonsemantically mediated routes are used in models of word processing. That verbal numerals mostly consist of more than one morpheme makes it difficult to form any comprehensive theory of verbal numeral recognition. Such an endeavor would require investigation into how morphologically complex words are recognized.

Arabic Numerals

Arabic numerals are of great importance to mathematical cognition literature. Arabic numerals utilize base 10 and uses place coding so that units, tens, hundreds, etc. are consistently held in the same position. This coding requires the use of the digit 0 in filling empty places (e.g. 408). Usefulness of Arabic notation can be seen when compared to usage of words in simple arithmetic problems. People are faster in such calculations if numerals are represented in digits rather than words (Campbell, 1994).

In general, research converges on the conclusion that small numbers in the Arabic system are recognized as a whole (Brysbaert, 2005). However, there is disagreement as to the definition of "small". Some believe that these small numbers are single digit numbers, while others extend the definition to include two-digit numbers as well. Another debate surrounding Arabic numerals is whether semantic activation is needed for the processing of Arabic numerals. More recent research on recognition of small Arabic numerals indicates that digits activate their meaning faster than words. In addition, they also require semantic mediation for further processing. This processing is more similar to picture recognition rather than word processing (Brysbaert, Fias, & Reynvoet, 2001; Fias, 2001). While meaning refers to the magnitude of the numeral, it can also activate encyclopedic and episodic knowledge associated with the numeral, especially if the numeral is used often in a nonquantitative way (e.g. Highway 99). Two-digit Arabic numerals sometimes require decomposition into tens and units and recognized as a combination of powers of 10 (McCloskey, 1992), but there is also evidence that they can be processed as one unit as a whole (Nuerk, Weger, & Willmes, 2001). A model was developed to explain these seemingly conflicting results in an assumption that both types of processing occur in parallel (Dehaene, 1992; Dehaene & Cohen, 1995). This model has Arabic numerals activating an analogue magnitude representation on the number line while also activating a visual Arabic number form that represents numbers as an internal string of digits. With this model, Dehaene suggests that there exists a lexicon for Arabic numerals similar to the orthographic lexicon for visual word recognition. This allows for extensive Arabic numeral processing before full activation of meaning. Arabic numerals of three digits or more, on the other hand, almost always require parsing, as evidenced by significant increases in responses latencies when compared to two-digit numerals (Brysbaert, 2005).

Spatial Representation

Spatial and number processing are closely related. Adults often use visuo-spatial representations of numbers (Seron et al., 1992), and increased visuo-spatial skills have been shown to be related to increased mathematical ability (Reuhkala, 2001), providing evidence for the transformation of number representations into visuo-spatial coordinates. Learning disorders (Dehaene & Cohen, 1997; Rourke & Conway, 1997) have also highlighted this relationship, as have brain imaging studies (Dehaene, Piazza, Pinel, & Cohen, 2003; Milner & Goodale, 1995) indicating activation of structures within the parietal lobes for tasks involving number processing or spatial transformations. Gobel, Walsh, and Rushworth (2001) further showed that transcranial magnetic stimulation (TMS) of the left and right parietal cortices leads to decreased performance in visuo-spatial search as well as in performing number comparisons.

While this provides evidence for a relationship between numbers and space, it does not specifically show that true numerical representations or processes are associated with spatial representations. This might instead be a result of shared support structures, such as the use of visuo-working memory in symbol manipulation during mental arithmetic tasks (Lee & Kang, 2002). Fischer (2005) does, however provide support for semantic representations of number magnitudes as being spatially defined and that they can further be conceptualized as positions on mental number line. This mental linear number line has been used to explain basic patterns in numerical cognition (Moyer & Landauer, 1967; Restle, 1970). It has also been supplemented by the idea that a hypothetical mental number line that also has spatial orientation within itself (Fischer & Fias, 2005). Fischer and Fias (2005) show this that this spatial cognitive representation of numbers is not fixed and that spatial number coding characteristics are heavily influenced by task-specific numerical and spatial parameters. Additionally, spatial coding of numbers is not under strategic control. Instead, this happens automatically.

Debate surrounding Mathematical Representation

Within math, representations are internal abstractions of mathematical ideas, and these abstractions are developed by learners through experience. There is an ongoing debate in cognitive psychology and mathematics education surrounding mental representations of math, or even mental numerical representations, in how the mind operates with representations. For example, do infants rely on an object-file representation or analog-magnitude representations quantity tasks? Evidence in favor of both an object-file representation (e.g. Corre & Carey, 2007) and analog-magnitude representations of number (e.g. Xu & Spelke, 2000) has been found. It is also possible that external representations can inform internal representations. There is debate stemming from this for what role symbols play in internal representations. For instance, to what degree are learners able to extract a mathematical concept form an external representation (e.g. graph, algebraic equation) or symbols? Similarly, there is also research exploring how much do learners glean from language aspects (e.g. metaphors) of mathematical concepts (e.g. Brannon, 2006).

Some argue that only numerals have exact lexical meanings for children, with the assumption that children cannot use a scalar implicature to inform numeral meaning. This claim is related to the finding that children similarly cannot exploy this ability for

quantifiers (Papafragou & Musolino, 2003). However, Barner & Bachrack (2010) argue against this view, instead hypothesizing that children's early interpretation of numerals does not utilize scalar implicature. Rather, they argue that the observed differences between numerals and quantifiers are a result of differences in availability of the related scales. They note that children can make scalar inferences in numeral interpretation, can assign approximate interpretations to numerals when first learning their meanings, and can also improve quantifier interpretations when there are scalar alternatives known to them.

It is possible that language does not inform our representations of number. Rather, exact number might be a cultural invention that allows us to keep track of cardinality of large sets across changes in modality. There is support for this shown from the study of indigenous groups whose languages have limited number words. These groups demonstrate approximate but not exact numerical ability (Gordon, 2004). Frank, Everett, Fedorenko & Gibson (2008), however, show that number words do not affect representations of number, using the Pirahã language as an example. The Pirahã has a uniquely limited numerical system. Their language has no word for number or exact quantity, even excluding the number "one". By use of matching tasks, Frank et al. (2008) showed that Pirahã speakers could identify exact matches for larger numbers of items but could not for for matching tasks requiring memory retrieval. Thus, despite lack of number words, Pirahã speakers' ability to perceive exact quantities was not reduced.

Early Development of Arithmetic Skills

Foundational elements of arithmetic begin before formal school instruction. Children already have developed mathematical cognition before starting school, and already have a rudimentary knowledge of arithmetic facts, procedures, and concepts. Early arithmetic knowledge allows us to study both general cognitive development as well as for mathematics specifically. (Geary, 1995; Newcombe, 2002). It also greatly informs educational assessment and instruction at the earliest and most critical stages of formal mathematics instruction. Instructional practices that do not match cognitive skills of the intended age group can be harmful to learning and achievement (Seo & Ginsburg, 2003). For ideal instruction and early assessment, developmental stages for children's knowledge of arithmetic must be taken into account (Ginsburg, Klein, & Starkey, 1998). Though children likely receive varying levels of direct and intentional instruction prior to school from teachers, family members, or other sources, it is at least reasonable to assume that a majority receive relatively little direction instruction in arithmetic prior to elementary school.

As seen earlier, infants already appear to be able to discriminate between small sets of objects that contain different numbers of elements. If infants possess this preverbal and pre-symbolic ability of quantification, then might they have a similar capacity for the arithmetic operations on quantities? Infant capability of arithmetic, at least with very small set sizes, was shown in several early studies (e.g. Cohen & Marks, 2002; Feigenson, Carey & Spelke, 2002; Wynn, 1992). However, these results have since been challenged (Baillargeon, 1995; Cohen & Marks, 2002; Haith & Benson, 1998). In a general argument (Simon, 1997), it is proposed that what appears to be numerical and arithmetical processing may instead arise from general-purpose, non-

numerical processes (e.g. the ability to distinguish individual entities). Simon suggested that infants create internal representations that could be object files (Kahneman, Treisman, & Gibbs, 1992) that initially contain little or no feature information. Only later is this file updated by attentional processes. Infants are also able to map an internally represented set to an external display with a one-to-one mapping. When the comparison process leads to a mismatch, expectation is violated, which leads to increased visual attention. This process might first appear to be numerical processing, but in truth might arise from attentional and memory processes needed for particular tasks. More recent studies provide evidence that is consistent with the view that non-numerical capacities are being used as infants perform small-number discriminations and arithmetic. In this general process, when presented with numerical tasks, infants are able to use basic perceptual and attentional mechanisms to individuate a small number of objects. They then represent them internal as object files. Attentional processes update this file with increasing level of detail for the corresponding object, including information about continuous extent. All this information is available in working memory for later processing and computation.

By contrast, older children do show some ability for solving arithmetic problems. This ability depends on task-dependent characteristics and presentation. Evidence from a variety of studies (Brannon, 2002; Starkey, 19922) support the conclusion that preschool children do indeed understand that addition increases set size and that subtraction decreases set size. These only shows sensitivity to such concepts but are not able to reveal how these children represent and process numerical information in these tasks. Preschool children also show some ability for exact addition and subtraction, with great variation of ability across age. Ability also depends on problem characteristics and presentation of problems (Huttenlocher, Jordan, & Levine, 1994; Starkey, 1992).

Most preschool children employ counting or retrieval in their solution procedures for arithmetic tasks (e.g. Siegler & Jenkins, 1989). Children's counting procedures do change with practice (Baroody, 1987; Groen & Resnick, 1977), and are sometimes supplemented by using external referents, such as fingers. With this ability to use counting in arithmetic, children might also be able to map objects or fingers to the number line, allowing them to count up and down the number line to add or subtract numbers. Retrieval, on the other hand, refers to rapid process of solving an arithmetic problem by accessing an exact answer from memory directly and bypassing any mental computation or counting (Ashcraft, 1982, 1992; Siegler, 1988).

How do preschool children choose among this variety of solution procedures for arithmetic problems? While this remains unclear, various computational models have been developed for this process (e.g. Shrager & Siegler, 1998; Siegler & Shipley, 1995) that have helped gain insight into why children retrieve answer successfully on some trials but not others. In general, these models followed patterns of children's behavior. Models would develop many new procedures, and from these, would choose counting-based procedures for difficult problems while choosing retrieval on more familiar problems. Both retrieval and accuracy improved with practice.

Representations in the Prefrontal Cortex

Transformations of equations required by algebraic problem solving involve retrieval of mathematical facts, which necessitates the distinctions between cognitive representation and retrieval. Some cognitive modeling has shown activity in the prefrontal cortex associated with retrieval required by algebraic manipulations and activity in the posterior parietal cortex associated with the transformations, yet there are similar activations in these regions for task manipulations (Anderson, Qin, Sohn, Stenger, & Carter, 2003; Qin, Carter, Silk, Stenger, Fissell, Goode, & Anderson, 2004). Danker and Anderson (2007) further attempt to distinguish activity in the parietal cortex and the prefrontal cortex by utilizing a multi-step algebra task, so that transformation is manipulated in an initial step and retrieval is manipulated in the following step. They employed two different cognitive models to test engagement levels of each of the parietal cortex and prefrontal cortex. Their first model assumed activation of only the parietal for transformations and only prefrontal for retrievals. When this model did not fit behavior and neural data, they designed a second model that allowed for the activation of both regions for each step in the task. Their surprising result was that both brain regions were active for each step but at differing levels of activation. This also highlights the ties between representation and retrieval in mathematical reasoning and also shows how cognitive modeling can be successfully used in understanding neuroimaging for task manipulations.

Number Processing in the Parietal Lobe

There is debate on whether the brain is predisposed to represent dedicated domains of knowledge. It has been proposed that the number domain is a possible biologically determined semantic domain (Dehaene, Dehaene-Lambertz, & Cohen, 1998). Certain attributes of domain specificity suggest that number and arithmetic are not merely cultural inventions and might have origins in brain evolution. One such attribute is that the ability to attend to numerosity and manipulate it internally in elementary computations present in animals (Hauser, Carey, & Hauser, 2000). Another is that a similar ability for elementary number processing is evidenced early in human development, before formal school or even the development of language skills (Spelke & Dehaene, 1999; Xu & Spelke, 2000), which suggests that numerical development is aligned with a developmental process that is based on mechanisms that have a long prior evolutionary history. Number processing may also depend on distinct neural circuitry. This can be identified consistently in subjects with many types of neuroimaging, neuropsychological and brain stimulation methods (Dehaene, Spelke, Pinel, Stanescu, & Tasivkin, 1999).

Use of parietal cortex in number processing was originally discovered from lesion data (e.g. Gerstmann, 1940), and further study revealed there is systematic activation of the parietal lobes along with precentral and prefrontal cortices during calculation. This was found using single photon emission tomography (SPECT) (Roland & Friberg, 1985) and was also supported with positron emission tomography (PET) (Dehaene, Tzourio, Frak, Raynaud, Coehn, Mehler, & Mazoyer, 1996; Pesenti, Thioux, Seron, & De Volder, 2000; Zago et al., 2001) and also functional magnetic resonance imaging (fMRI) (Burbaud, Camus, Guehl, Bioulac, Caille, & Allard, 1999; Rueckert et al., 1996).

Dehaene & Cohen (1995) further proposed that the parietal lobe contributes to the representation of numerical quantity on a mental number line. More recent studies have shown that mental arithmetic relies on a set of processes, and that many of these processes are likely not specific to the number domain. For example, Spelke & Tsivkin (2001) conducted studies of language interference in normal subjects that suggest that language based processes are crucial in exact but not in approximate calculation. Similarly, Lee and Kang (2002) show that attempting a spatial task interferes with subtraction but not with multiplication. Both these suggest that the neural bases of calculation is heterogeneous.

Theory for number processing includes a triple-code model, which predicts that three distinct systems of representation maybe be called upon depending on the task. These three systems are a nonverbal semantic representation of the size and distance relations between numbers (quantity system), a system for which numerals are represented lexically, phonologically, and syntactically (verbal system), and a system in which numbers can be encoded as strings of Arabic numerals (visual system) (Dehaene, 1992; Dehaene & Cohen, 1995). Dehaene & Cohen first suggested that parietal activations for number processing were the result of the contribution of the quantity system. However, they later noted that the left perisylvian language network clearly extends into the inferior parietal lobe, and that the posterior superior parietal lobes are strongly engaged in visual attention processes that may contribute to the visual processing of numbers. From this they see necessity in distinguishing which activation sites within the parietal lobe are associated with a semantic representation of numerical quantity and which correspond to nonspecific verbal or visual/attentional systems.

In further work, Hubbard, Piazza, Pinel and Dehaene (2005) propose that several circuits coexist in the parietal lobe and that these circuits account for most of the observed differences between arithmetic tasks. The first of these circuits is a bilateral intraparietal system associated with a core quantity system, the second a region of the left angular gyrus associated with verbal processing of numbers, and the third a posterior superior parietal system of spatial and non-spatial attention. For the purpose of this paper, I will cover only the bilateral horizontal segment of the intraparietal sulcus and quantity processing.

Dehaene, Piazza, Pinel and Cohen (2005) found that the horizontal segment of the intraparietal sulcus (HIPS) is a significant site of activation in neuroimaging studies of number processing, and with images of axial and sagittal slices show that this region is at the intersection of the activations observed in many different number processing tasks. The commonality of these tasks is the need to access a semantic representation of a quantity that a number represents. From this they propose that nonverbal representation of numerical quantity is present in the HIPS of both hemispheres and could be analogous to a spatial map or number line. This would support our seeming ability to know what a given numerical size means as well as know the proximity relations between numbers.

The HIPS is active whenever an arithmetic operation that needs access to a quantitative representation of numbers is needed for mental arithmetic. For instance, when subjects calculate, the HIPS was more active than when subjects simply read numerical symbols (Burbaud et al., 1999; Pesenti et al., 2000). This suggests that the HIPS is important in the semantic manipulation of numbers. When subjects have to

compute two addition or subtraction operations instead of one, activation increases (Menon, Rivera, White, Glover & Reiss, 2000). Also, the HIPS is more active when subjects estimate results of additions problems than when they compute exact solutions (Dehaene, Spelke, Stanescu, Pinel & Tsivkin, 1999). Subtraction appears to lead to greater activation than does multiplication (Chochon et al., 1999; Lee, 2000). Multiplication tables and small exact addition facts can be stored in rote verbal memory and thus require minimal quantity manipulation. However, some subtraction problems can be stored in verbal memory, yet are not typically learned by rote. Thus this requires actual quantity manipulations.

The HIPS also activates for number comparisons. This occurs whenever a comparative operation requires a numerical scale. For example, it is more active when comparing the magnitudes of two numbers than it is when only reading these numbers (Chochon et al., 1999). This systematic activation has been shown in a variety of paradigms using tomographic imaging (Pisenti et al., 2000; Pinel, Dehaene, Riviere, & LeBihan, 2001; Thioux, Pesenti, Costes, De Volder, & Seron, 2002) and also scalp recordings of event related potentials (Dehaene, 1996). This parietal activation in number comparison is typically greater in the right hemisphere than it is in the left hemisphere (Chochon et al., 1999; Dehaene, 1996; Pinel et al., 2001). Other conclude that this might indicate a possible right-hemispheric advantage in comparison as well as in other tasks that require an abstraction of numerical relations (Langdon & Warrington, 1997; Rosselli & Ardila, 1989). But even though this asymmetry exists, parietal activation is always present in both hemispheres in split-brain patients (Cohen & Dehaene, 1996; Seymour, Reuter-Lorenz, & Gazzaniga, 1994).

This activation of the HIPS for number comparisons is greater than activation that occurs when comparing other categories of objects that are non-numerical (Le Clec'H et al., 2000; Pesenti et al., 2000; Thioux et al., 2002). In addition, event-related potentials have shown more parietal activation for numbers than for other categories of words (Dehaene, 1995). Thioux et al. (2002) tested the specificity of the HIPS for the numerical domain for several tasks. Participants were presented with number words and names of animals that were matched for length. The HIPS revealed more activation for numbers than for animal names, both for comparison tasks and categorization tasks. This suggests that the HIPS shows category-specificity independent of task context.

Parametric studies further reveal that the HIPS activation is controlled in part by semantic parameters (e.g. absolute magnitude of numbers and their values relative to a reference point), suggesting that intrapareital activity is greater during operations with large numbers than with small numbers (Kiefer & Dehaene, 1997; Stanescu-Cosson et al., 2000). It is also controlled by numerical distance between numbers in a comparison task (Dehaene, 1996; Pinel et al., 2001). However, HIPS activation is does not depend on the particular modality of input in conveying numbers. Arabic numerals, spelled-out number words, and even nonsymbolic stimuli can activate the HIPS (Le Clec'H et al., 2000; Piazza, Mechelli, Butterworth, & Price, 2002; Pinel et al., 2001). In a study by Piazza et al. (2002), subjects were asked to attend to either the numerosity or to the color of series of visual events. The right HIPS was active when subjects attended to number. Subjects were also asked to attend to either numerosity or pitch in a serious of auditory stimuli. The right HIPS was similarly active when subjects attended to number. Thus

modality of the stimuli did not affect HIPS activation (Piazza et al., 2002). A study by Pinel et al. (2001) showed that activation of the bilateral HHIPS was found to correlate directly with the numerical distance between two numbers in a comparison task. Numbers were presented as either words or as digits, and results showed equivalent activation for either. Thus parametric studies appear to be consistent with the hypothesis that the HIPS codes abstract quantity meaning of numbers rather than the numerical symbols themselves.

Interestingly, HIPS activation can occur even when a subject is not aware of having seen a number symbol (Dehaene et al., 1998; Naccache & Dehaene, 2001). For these experiments, subjects compared target numbers to a fixed reference of 5. Before presenting the target number to subjects, another number was presented subliminally. FMRI show activation in the left and right intrapaietal regions when there was unconscious repetition of the same number. When the prime and the target matched, there was less parietal activation than when the prime and target did not match. This indicates that this region comprises distinct neural assemblies for different numerical quantities. More activation can be observed when there are two neural assemblies being activated than when only one assembly is activated.

Data gathered from neuroimaging on the HIPS thus suggests that this area is crucial for semantic representation of numbers as quantities. This could be functioning as our foundation of numerical intuition and immediate, and sometimes unconscious, understanding of a quantity in relation to others (Dehaene, 1992, 1997). This is true whether or not it is appropriate to a given task.

Learning Disabilities

There have been several studies over the past two decades that provide us with an initial understanding of the number, counting, and arithmetical competencies and deficits of children with mathematical learning disorders (e.g. Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2011; Desoete, Roeyers, & De Clercq, 2004; Geary, Hamson, & Hoard, 2000). Many children with mathematical learning disorders have an underdeveloped conceptual understanding of some aspects of counting (e.g. Geary et al., 2000). They understand inherent counting rules but commit errors on tasks that measure order or adjacency (Gelman & Gallistel, 1978). Poor counting knowledge in turn contributes to delayed ability in the use of counting in solving arithmetic problems and can result in deficiencies in detecting and correcting counting errors (Ohlsson & Rees, 1991). When trying to solve simple arithmetic problems and simple word problems, children with mathematical learning disorders can use the same types of strategies as normal children. However, their strategy mix and pattern of developmental change can vary (Geary, 1990; Hanich Jordan, Kaplan, & Dick, 2001; Jordan, Hanich, & Kaplan, 2003). Children with mathematical learning disabilities tend to rely on developmentally immature strategies, commit frequent counting errors, use lower-level counting procedures, and have difficulty in retrieving basic facts from long-term memory (Geary & Brown, 1991; Hanich et al., 2001; Jordan & Montani, 1997).

Children with mathematical learning disorders consistently differ from their normal peers simply in the ability to use retrieval-based processes in trying to solve simple arithmetic and word problems (e.g. Hanich et al., 2001; Jordan et al., 2003;

Temple & Sherwood, 2002). It seems that the ability to retrieve basic facts does not dramatically improve across elementary school years. This deficit in retrieval ability might result from difficulty in representing phonetic and semantic information in long-term memory (Geary, 1993) or might result from a deficit in ability to inhibit irrelevant information from entering working memory during problem solving (Barrouillet, Fayol, & Lathuliere., 1997; Conway & Engle, 1994; Geary et al., 2000).

Procedural deficits might stem from poor conceptual understanding of counting concepts (Geary et al., 2000) or poor working memory and central executive resources (Hitch & Mcauley, 1991). Other work indicates a reliable relation between accuracy of procedural execution and performance on a measure of working memory and central executive functioning (Geary, Hoard, Byrd-Craven & DeSoto, 2004). These provide evidence for confirming the relation between conceptual knowledge and procedural memory.

Children with deficits in ability to retrieve information from long-term memory also contributes to arithmetic fact retrieval deficits of children, who also show deficits on measures that assess skill at accessing other types of semantic information (Geary, 1993). Some children with mathematics learning disabilities that have fact-retrieval deficits do have a language representation deficit as well (Geary et al., 2000), while others may not share this deficit (Jordan et al., 2003). Problems with spatial ability are also associated with misalignment in setting up arithmetic problems, for example when writing out an addition problem vertically or horizontally. This also affects understanding of Arabic numerals that rely on a positional, base-10 system.

Thus, most children with mathematical learning disorders have near-normal number-processing skills at least for simple numbers. However, they are lacking in representational and processing skills for larger numbers. These children have persistent deficits in some areas of arithmetic and counting knowledge, and also possess lower level understanding of some counting principles when compared to similar-aged children without mathematical learning disorders. In fact, they often use problem solving procedures in arithmetic that are used more often by younger children, and commit frequent procedural errors. Procedural skills for simple arithmetic has been shown to improve over elementary school years, which provides promise that such an early deficit may not be a permanent cognitive disability. Many children with mathematics learning disorders also have difficulties retrieving basic facts from long-term memory, however, and this deficit may not improve as much as procedural skills over time. Some developmental delays and deficits stem from a combination of disrupted functions of the central executive, including attentional control and weak inhibition of irrelevant information, or difficulty with representation and manipulation within language systems. Mathematics learning disorders can theoretically also result from a less developed visuo-spatial system.

Generally, the relation between deficits in simple calculation or number processing and other numerical or non-numerical skills has not full been addressed. For example, it is not clear if treatment of one particular deficit affects other aspects of the number-processing domain. Furthermore, how additional cognitive deficits affect treatment of calculation abilities has yet to be examined systematically.

Cognitive Engineering Applications in Mathematics Education

Cognitive Tutors

Cognitive models of mathematics learning are highly useful in designing and evaluating school curricula and in designing software aimed to supplement classroom learning. Ritter, Anderson, Koedinger and Corbett (2007) developed one such model to create a "cognitive tutor". This project began with Anderson's development of the adaptive control of thought (ACT-R) theory of cognition. This theory sought to outline a unified theory of cognition and was implemented as a computer program. Initial results in learning, memory, and problem solving were promising (Anderson, 1983), but were limited to laboratory settings. The goal of ACT-R is to interpret behavior, and to accomplish this, the models must represent knowledge accurately and also know that knowledge results in associated behaviors. In particular for education, this leads to predictions for what students can do as well as predictions for what learning and study activities will help students succeed in achieving curricular goals. Cognitive models such as Anderson's were later turned into tutoring systems called "cognitive tutors". Two such cognitive tutors for geometry proofs (GPT, ANGLE) were shown to be effective in a school environment (Koedinger & Anderson, 1993), though success depended on teacher skills in incorporating this tutor into long-term curricular goals. Subsequent tutors were designed in tandem with a complete course that included training for teachers, and this new curricula showed educational effectiveness (Koedinger, Anderson, Hadley & Mark, 1997; Koedinger, Corbett, Ritter & Shapiro, 2000). From this success, Ritter, Anderson, Koedinger and Corbett (2007) saw need to include four essential components in the development of a research-based curriculum. These components are: basing the curriculum on a solid theoretical foundation, applying basic theory to the particular domain and objectives of interest, evaluating results, and developing and implementing a methodology for improving the curriculum on the basis of use. I will discuss each of these components in detail below.

ACT-R acts as a theoretical basis for cognitive tutors (Anderson, 1990, 1993; Anderson, bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Anderson, Lebiere, Lovett, & Reder, 1998). This theory has sought to model characteristics of human behavior. These targeted characteristics include error patterns and response times over a variety of cognitive tasks. Studies involving ACT-R have been conducted both in laboratory settings and classroom settings. In the case of education, Ritter et al. (2007) propose three fundamentals. The first is that procedural and declarative knowledge should be distinguished, and that all tasks involve a combination of both these types of knowledge. Procedural knowledge tends to be more fluent and automatic, while declarative tends to be more flexible and more widely applied than procedural knowledge. A second fundamental is that knowledge required for complex tasks are comprised of the declarative and procedural knowledge components for that task. The third fundamental is that declarative and procedural knowledge both improve with use. "Strong" knowledge can be recalled to attention very quickly and with certainty, while retrieval of "weak" knowledge can be difficult and time consuming if it is indeed possible at all. ACT-R embraces a gradual process of encoding, strengthening, ad proceduralizing knowledge. Interaction between declarative and procedural knowledge results in active engagement of concepts underlying procedures. This allows students to generalize knowledge (Rittle-

Johnson & Koedinger, 2002, 2005; Rittle-Johnson & Siegler, 1998; Rittle-Johnson, Siegler & Alibali, 2001). Thus education activities require structure surrounding practice of procedures.

ACT-R provides a foundational theory for cognitive modeling but does not highlight skills necessary for specific tasks. Much applied research in mathematics has addressed this shortcoming in identifying skills and methods that students use to complete specific math tasks (Corbett, McLaughlin, Scarpinatto, & Hadley, 2000; Koedinger & Anderson, 1990). Besides aiding design of math tasks, ACT-R theory does guide instruction in cognitive tutors owing to use of an active cognitive model (Corbett, Koedinger, & Anderson, 1997). This allows the model to follow student actions in determining student strategy in solving a problem. Additionally, each student action is associated with particular skills that act as references to knowledge components in the cognitive model. Individual student performance can be tracked for these skills over time, and this is used to select problems that emphasize the skills that students display weak knowledge. This skill model is further used to enact mastery learning. Once skills in a section are deemed up to the level of sufficient mastery, the student can move on to the next section of the curriculum.

Careful evaluation is crucial in the difficult task of curriculum development. Such evaluations have been conducted over individual units of instruction and entire curricula (e.g. Aleven & Koedinger, 2002; Anderson, Corbett, Koedinger, & Pelletier, 1995) and have been considered in tandem with performance on standardized tests (Koedinger et al., 1997; Morgan & Ritter, 2002). Students who used a cognitive tutor integrated into their curriculum performed better on standardized tests, most notable on tests of problem solving and multiple representations.

ACT-R also acts as a guideline for educational pedagogy, and aids in testing and improving curriculum (Ritter et al., 2007). A cognitive tutor can observe students over varying timescales, from10-second intervals to the span of an academic school year. It can evaluate student behavior and predict what students knows and to what degree of mastery the student has for any given skill. Student errors should theoretically decrease over time, and as this happens, cognitive tutors will introduce higher difficulty tasks. While the cognitive models employed are not without error, steps have been taken to investigate flaws in these models (Cen, Koedinger, & Junker, 2005; Junker Koedinger & Trottini, 2000).

These ACT-R based cognitive tutors have led to an impressive amount of collected data, which promises to lead to the detection of increasingly subtle factors that influence classroom learning, and furthermore inform our knowledge of students' mathematical cognition. This can in turn aid in structuring education to more effectively help students learn mathematics.

Online Tools for Mathematics Education

Software and online tools for learning math have become increasingly common in recent years. One such tool is the Virtual Math Teams (VMT) environment (Stahl, Rose, O'Hara, & Powell, 2010), created to support dynamic mathematics in education. VMT employs software agents that interact synchronously with small groups of students working in an online environment. These agents encourage academically driven

interactions and facilitate in keeping students on task. The environment includes a social-networking portal, an application that integrates synchronous text chat, social awareness indicators, and an asynchronous community wiki.

Educational theory backs the VMT environment. The VMT requires students to develop skills in mathematical conversations (Sfard, 2008; Stahl, Rose & Goggins, 2010). Simply using any online resource allows students to engage in conversations, but this unguided practice typically is less practical and efficient with fewer educational benefits (Stahl, 2006). The VMT provides structure and tools needed for mathematical discussions and aids educators in developing math topics for classroom exploration and discussion.

To accomplish productive student focus on math topics, the VMT employs scaffolding or scripting of discussion (Kobbe, Weinberger, Dillenbourg, Harrer, Hamalainen, Hakkinen, & Fischer, 2007). Even a well-designed topic can be derailed after a starting in the correct thread of thought in discussion. Furthermore, the VMT provides feedback to aid student understanding. This feedback can come from teachers or other student groups working on similar projects. The VMT can also guide collaborative processes synchronously. This helps a great deal when a teacher must otherwise keep track of multiple groups at once. The level of support offered by the VMT allows for math discussions that are not typical to most classrooms (Boaler, 2008). Even existing traditional classrooms can use this environment to supplement the classroom experience with a variety of adaptable and personalizable interactions (Scher, 2002).

Already there has been support for the effectiveness of virtual math teams (Cui, Kumar, Chaudhuri, Gweon, & Rose, 2009) in pre and post-test learning gains. Online tools such as this show great promise for developing tools for aiding student learning of mathematics.

Conclusion

We have seen that mathematical cognition includes a wide array of theoretical and practical applications. These diverse areas nonetheless together provide mutually support in informing developments across areas. The multidisciplinary nature of mathematical cognition emphasizes the importance of understanding mathematical cognition within many domains, from industry to education, and highlights new issues surrounding mathematical skills with implications across research and education.

Chapter 3: Math Anxiety and Psychological Stress

(Note: Work presented here originally appeared in the Proceedings of the 35th Annual Meeting of the Cognitive Science Society and in Frontiers in Psychology 6 (2015))

Math Anxiety

Math anxiety is usually seen as a feeling of apprehension or fear when considering any math task. Richardson and Suinn (1972) created the first instrument for measuring math anxiety in 1972, called the Mathematics Anxiety Rating Scale (MARS). For this survey, participants rate their levels of anxiety (using a Likert scale) for 98 various math-related tasks and situations, such as taking a math test or calculating the tip for a restaurant bill. Shorter variants have since been formulated, such as the abbreviated MARS (Alexander and Martray, 1989; Plake & Parker, 1982) which uses a selection of 25 questions from the original MARS. Scores for these shorter versions correlate with scores on the original test and show excellent reliability. These versions are now more typically used in recent research rather than the original 98-question version.

Math anxiety can carry both educational and personal repercussions, and the most common consequence is simply math avoidance (e.g. Dew, Galassi, & Galassi, 1984; Ashcraft, 2002). People who have high math anxiety tend to avoid math. Those with high math anxiety also carry negative views of math and negative views of their own math ability (e.g. Eccles, 1986; Ashcraft, 2002). There also exist negative correlations between math anxiety and motivation and self-confidence in math (Kloosterman, 1988). Resulting math avoidance can take the form of avoiding enrolment in math classes, avoiding math lectures, or avoiding math homework. This can start as early as elementary school and can escalate through high school and college class choices, and later can lead to avoidance of careers that involve a significant amount of math.

Surprisingly, math anxiety has only a weak relation with overall intelligence (Hopko, Crittendon, Grant, & Wilson, 2005), and this is likely exaggerated by the math-based quantitative items that appear on IQ tests (Ashcraft, 2002). People with math anxiety will likely perform worse on these tasks than those with low math anxiety. Typically, it is believed that women have higher math anxiety than men. However, this difference has been shrinking in recent years so as to be virtually non-existent (e.g. Tapia & Marsh, 2004). However, some groups may display greater gender difference in math anxiety simply because of willingness to report math anxious tendencies in self-report surveys such as MARS. Also, those with high math anxiety also tend to have high overall anxiety.

Effects on Math Ability

Avoidance of math will ultimately lead to lower ability in math. If students avoid math, they will lack the practice and background to succeed in math-related tasks and this can lead to lower educational achievement, such as in standardized tests (Fennema, 1989). This poses a question for the underlying reasons for low performing students. If a student with math anxiety does not do well on a test, is this due to low competence or due to math anxiety directly? If the relationship between anxiety and competence is constant at all levels of ability, then this poses an unsolvable problem, as differences in competence will make it impossible to evaluate math ability for differing levels of math

anxiety. There are ways to delve into this problem. Faust, Ashcraft, and Fleck (1996) suggested testing additional samples of participants on various mediums and under different conditions. For example, participants might take a paper-and-pencil version of math problems without time restriction. They found no anxiety effects for whole-number arithmetic problems using such a test format. However, when participants took a timed online test, there were significant anxiety effects on the same problems. In another approach, a standard math achievement test was given to low, medium, and high math anxiety students (Ashcraft, Kirk, & Hopko, 1998). They found that test scores decreased with higher levels of math anxiety. They also examined scores as the test progressed from easier to more difficult items. In the first half of the test, which consisted of whole number arithmetic problems, students did not display math anxiety effects. However, the second half of the test, which included mixed fractions, percentages, questions with unknowns, and factoring, showed strong anxiety effects. For this second half the test, there was a negative relationship between math anxiety and performance. This showed that people with high math anxiety do not necessarily have an overall deficit in math competence. They can at least show equal ability as their low-anxiety counterparts in performing whole number arithmetic problems. However, higher level math tasks show that math anxiety does affect competence.

But how robust is the relationship between math anxiety and ability? Hembree (1990) introduced treatments for student with high math anxiety. These treatments effectively increased student scores to the same level as students with low math anxiety. Note that these treatments were not based on teaching or math practice, and so the resulting increase in scores was not because of a direct increase in competence via these treatments. Hembree instead suggests that students' original low scores resulted from the presence of math anxiety. Once their math anxiety was alleviated, students were able to perform closer to their true level of ability.

Effects on Cognitive Ability

In early studies, Ashcraft and Faust sought to investigate whether math anxiety had an impact on cognitive processing during problem solving. They found that math anxiety had little influence on math performance with single-digit addition and multiplication (Ashcraft & Faust 1994; Faust et al., 1996). But it did become an influence when problems involved decision-making the revolved around "number sense" (Dehaene, 1997). When participants were asked to make true and false judgements, those with high math anxiety made more errors when problems became increasingly implausible (e.g. 4 + 11 = 50). But those with low levels of math anxiety made fewer such errors on implausible problems. They found further anxiety effects for arithmetic problems that involved larger numbers (e.g. two-digit multiplication). Those with high anxiety answered questions almost equally rapidly as those with low math anxiety, but with many more errors. On some level, this behavior might reflect general avoidance behavior of those with high math anxiety. But their lower accuracy might also stem from reduced time spent practicing in math classes. Just as they are quick to provide any answer, it is likely they enact similar behaviors for any of their math classes in rushing through their homework or other class tasks. Further results showed that addition problems that required carrying also posed more difficulty for high math anxiety

participants. The increase from no-carry to carry problems for highly math anxious participants was 753ms, compared to the 253ms difference for low math anxiety participants. This significant difference does not even take into account the higher accuracy of the low math anxiety participants. It appears that carrying, a procedural part of arithmetic, placed higher demand on working memory, which is used for conscious mental processing. From this, it can be reasoned that math anxiety affects cognitive operations that utilize working memory.

Ashcraft and Kirk (2001) explored this idea by testing one and two column addition problems, half of which required a carrying operation. They included this test in a dual-task procedure, having participants perform mental math as the primary task while trying to remember random letters as a second task that used working memory. Two or six letters were displayed to each participant before each mental addition problem. After the participant provided an answer to the math problem, they tried to recall the letters in order. Ashcraft and Kirk hypothesized that, if the math task depended on working memory, performance on the primary task of mental addition would decline with increasing numbers of letters that had to be memorized. This would be a sign that the two tasks in combination would have over-taxed the limited capacity of working memory. In their results, addition problems that required carrying resulted in lower accuracy more for high math anxiety participants than for low math anxiety participants. This was most pronounced for difficult secondary tasks that required participants to memorize six letters. For problems that required carrying, those with high math anxiety had an accuracy rate of 60% in the high-load condition, while those with low math anxiety had an accuracy rate of 80% for the same high-load condition. In the light-load condition, both groups had an accuracy rate of 88%. They conducted a similar experiment that replaced the math task with a verbal one, and found no differences between anxiety groups. Differences only arose between anxiety groups when using an arithmetic task.

Eysenck and Calvo (1992) have presented similar findings. They created a model of general anxiety effects. In this model, called processing efficiency theory, general anxiety is hypothesized to derail current working memory processes. They reason that those with general anxiety use more attention to their intrusive thoughts and worries that interfere with a current task. In considering math anxiety, similar disruptive thoughts probably evoke someone's low self-confidence in math or fear of math. Math anxiety can lower math performance by giving primary attention to these thoughts, which act as a secondary task that lowers the attention given to the math task. Thus, our dependence on working memory to complete a math task results in a corresponding drop in cognitive performance. Spada, Nikcevic, Moneta, & Wells (2007) further showed that perceived general stress can negatively affect metacognition and invoke negative emotion. They also found positive correlations between perceived stress, anxiety, and depression.

Not all math tasks require such a heavy demand on working memory. Math facts learned through rote memorization (e.g. the single digit multiplication table) require little working memory. For tasks such as these, math anxiety makes no noticeable difference in performance. However, more complicated problems that required more layered procedures and operations (e.g. carrying, long division) place a heavier load on working memory and these tasks are therefore impacted by math anxiety. It is plausible that this

pattern of increasing demand on working memory will escalate with the difficulty of math problem. For example, an algebra or calculus problem likely draws even more from working memory, and this increased load might results in increased performance effects derived from math anxiety.

Causes of Math Anxiety

Math anxiety is indeed considered a specific type of anxiety or phobia. It can have both short-term cognitive consequences and long-term impact on educational and career achievement. While there is still on-going empirical work for the causes of math anxiety, there is no cohesive collection of research. Some examples of possible causes are family influence (e.g. Ferry, Fouad, & Smith, 2000; Stuart, 2000), classroom environment (e.g. Geist, 2010; Jackson & Leffingwell, 1999), and stereotype threat. Stereotype can take many forms, including gender (e.g. Spencer, Steele, & Quinn, 1999; Aronson, Lustina, Good, Keough, Steele, & Brown, 1999) and race (e.g. Osborne, 2001). One difficulty in studying cognitive consequences of math anxiety is that accurate assessment of actual math ability is difficult to achieve. We saw that math anxiety interferes with one's ability to perform complex math tasks, and it follows that performance will also suffer from this interference. While the underlying mechanisms for this are not yet clear, it is possible that, while intrusive thoughts are a tempting scapegoat, it could be that math anxious people are not able to inhibit these thoughts (Hopko, Ashcraft, Gute, Ruggiero, & Lewis, 1998).

More is being discovered about math cognition through the study of cognitive neuroscience. Notably, Dehaene, Spelke, Pine, Stanescu, and Tsivkin (1999) found strong evidence that math intuition is linked to both linguistic competence and visuo-spatial representations through brain imaging. For instance, exact arithmetic can be processed in a language-specific format, but this does not hold true for novel facts. However, approximate arithmetic is independent of language. Instead, it employs numerical magnitudes and the bilateral areas of the parietal lobes that process visuo-spatial tasks. Instead, networks that use word-association processes are employed. This is just a start for the intricate combinations of systems needed for math cognition. Math anxiety will have its own place within neural activity that characterizes this particular phobia. Young, Wu, and Menon (2012) have shown that math anxiety in 7- to 9-year-old children was associated with activation of right amygdala regions typically used for processing negative emotions. Furthermore, math anxiety was also linked with reduced activation in the posterior parietal and dorsolateral prefrontal cortex regions, which are typically used in math reasoning tasks. These effects associated with math anxiety were independent of general anxiety, intelligence, working memory, or reading ability. Lyons and Beilock (2012) similarly showed math anxiety to be linked to pain network activation when subjects are anticipating a math task. They debate whether math anxiety is only psychological if it is triggered through a visceral threat or even an actual experience of this pain. Their results suggest that while the anticipation of math results in pain network activation, the actual math task itself does not. These initial studies provide possible mechanisms for why people with high math anxiety tend to avoid math, and why math anxiety is often associated with lower math ability.

Measuring Psychological Stress

Questionnaires are by far the simplest method for measuring stress. Surveys, such as the Perceived Stress Survey, however, measure just that – perceived stress. For example, does the Math Anxiety Ratings Scale (MARS) directly measure math anxiety, or merely people's perception of their stress response to math? People might further report answers they feel are expected of them. Math anxiety is widely known and acknowledged, and some may report feelings of math anxiety because they believe this to be typical of their peers, even if they do not experience true math anxiety. However, stress can also be reliably be measured through biological production of cortisol.

Cortisol is a hormone triggered by stress responses in humans. It is often used as a biomarker of psychological stress, as well as related mental or physical diseases. Cortisol is not affected exclusively by major events. Daily occurrences of stress also affect cortisol levels and associated health consequences, though less is known about the effects of daily chronic stress through minor daily events. Van Eck, Berkhof, Nicolson, & Sulon (1996) explored the neuroendocrine effects from minor daily events. In their study, they measured perceived stress, traits, mood states, and stressful daily events, and salivary cortisol. They used Perceived Stress Scale scores to select "high stress" and "low stress" subjects from a pool of male participants from a similar working class. Choosing males for cortisol research is common, as they have fewer hormonal fluctuations that affect cortisol levels independently of stress-inducing events. All participants completed an Experience Sampling self-report. Saliva samples were collected 10 times a day over a span of 5 consecutive days. Part of their results showed that trait anxiety and depression, but not perceived stress, were tied to modest cortisol elevation. Interestingly, no effects were found for recent life events, chronic difficulties, trait anger, or psychosomatic symptoms. Participant distress was measured by Negative Affect and Agitation mood states, which were linked with higher cortisol levels. Positive Affect had no effect. Daily stressful events were associated with higher cortisol. Perceived stress, depression, and anxiety did not affect cortisol levels, but mood appeared to moderate the association between stressful daily events and cortisol levels. In general, their results indicate that negative affectivity is both a confounder and leads to increased cortisol production during normal daily activities. Even small changes in mood states are associated with higher cortisol production, and this might point to a link between subjective stress and health outcomes.

Limitations

Associations between cortisol secretion and responses on subjective stress tests tend to be weak (Hellhammer, Wüst, & Kudielka, 2009). This is not surprising, as self-report measures measure perceived stress, and tests can be based on varied theoretical constructs. Gender and personality traits further confound measures from self-reported stress. The low association may also arise from neuroendocrine factors. Psychobiological processes that activate the hypothalamus-pituitary-adrenal axis (HPAA) can also be measured through cortisol. Thus activations of HPAA reactivity, such as the hippocampus, hypothalamus, pituitary, and adrenals, can influence cortisol measures. While salivary cortisol is a valuable tool in measuring stress, it is necessary to acknowledge that there can be significant sources of variance.

Methodology

Methodological issues arose early in cortisol response research. Before the 1980s, research on acute and chronic effects of stress used cortisol levels as a measure of individual stress responses (Kirschbaum & Hellhammer, 2009). Original measures of cortisol for early research were obtained from blood samples, an invasive and costly method, which might be a confound in itself. One can imagine that even the act of a blood draw would lead to increased levels of HPAA activation. In the 1980s, a biochemical assay was developed that could measure cortisol in the lower nanomolar range from salivary samples. Though reliability was initially questioned, this methodology proved to be as equally reliable as blood collection methods (Kirschbaum & Hellhammer, 1989). Cortisol levels measured from saliva samples also correspond well with measurements taken from blood samples (Kirschbaum & Hellhammer, 2009). Salivary cortisol measurement has become the most common choice for field studies or studies conducted in natural environments. This methodology offers several advantages over cortisol measured from blood. It is less invasive, less costly, and takes less time to collect. It does not require collection to take place in a laboratory, and medical training is not necessary for those collecting samples. Samples can also be taken over short intervals without undue stress from multiple blood draws and is safe for use with children. Collection materials, typically a salivette, are relatively inexpensive. A salivette is simply a polypropylene tube that contains an inner centrifuge tube and cotton swab. Subjects place this cotton swab in their mouth and under their tongue for 2 minutes, and saliva is absorbed into the cotton swab. The swab is then placed back into the tube. Subjects can also salivate directly into small containers. Cortisol is very stable in saliva, so it can remain at room temperature several weeks without risk to the cortisol levels present. This allows samples to be stored easily, or even mailed to other locations. Samples can be kept even longer if frozen. There are several commercial assay kits that can be used for the biochemical analysis of salivary cortisol. The most widely used assays used in laboratories is an immunoassay with radioactive and other tracers. Smaller scale labs, such as those in university departments, often use an enzyme-linked immunosorbent assay (ELISA) that uses a horseradish perozidase labelled cortisol tracers (Cooper, Trunkfield, Zanella, & Booth, 1989).

Both blood and salivary cortisol levels follow a circadian rhythm of daylight experience throughout the span of a day. Cortisol reaches a peak level early in the morning. While it then follows a subsequent decreasing trend, there are a few smaller spikes that occur throughout the day, even without the introduction of a psychological or physical stressor. Some of these spikes are caused by meals throughout the day. However, cortisol shows no large increases during the afternoon and evening. This circadian pattern is quite robust, and remains stable over months of observation.

When subjects are introduced to situation meant to induce stress (a stressor task), cortisol does not immediately show a strong response. Instead, it takes several minutes before HPAA response becomes apparent. Adrenocorticotropin (ACTH) levels will increase within 5 minutes of the onset of s stressor. After stimulation from ACTH, cortisol is synthesized and secreted from adrenal glands. For shorter stressors, cortisol will peak soon after the end of the stressor. Stressors induced over a long period of time can result in increased cortisol production over a similarly extended period of time. For

example, a 3-hr stressor task may result in a cortisol peak 20 minutes after the end of the stressor.

Applications

Measures of salivary cortisol have been used show the cortisol production response to a wide array of stressors, such as public speaking (Bassett, Marshall, & Spillane, 1987), marathons (Cook, Read, Walker, Harris, & Riad-Fahmy 1986b), thriller movies (Hubert & de Jong-Meyer, 1989), and academic exams (Hellhammer Hubert, & Schurmeyer, 1985; Jones, Copolov, & Outch, 1986). It has become accepted that psychological stress can increase activation of the HPAA. This is especially true of situations that have a high level of ego-involvement, novel situations, low control, and low predictability. For these situations, corticotropin releasing hormone (CRH) and ACTH are released, which causes an increase in cortisol production. Notably, using salivary cortisol collection methods has advanced studies for children and infants, as venipuncture is not a safe or ethical option for such young subjects. Gunnar, Connors, and Isensee (1989) studied cortisol production response for newborns who underwent 2 hospital discharge examinations. During the first exam, cortisol showed a large increase, while cortisol response to the second exam was far lower. Spangler and Scheubeck (1993) found similar stress responses in newborns to discharge examinations. However, they also observed that infants reacted to both exams only when they assessed heart rates with a heart rate monitor attached to the child's chest simultaneously. Similar increases in salivary cortisol was found through inoculation of infants (Lewis & Thomas, 1990) and brief maternal separation (Larson, Gunnar, & Hertsgaard, 1991). These results suggest that the adrenocortex is responsive from infancy, and as early as the postnatal period. In addition, these response patterns largely mimic those of adults.

Much research has also been devoted to the role of cortisol in mental and physical illness. Increased cortisol is associated with human aging (e.g. O'Brien, Lloyd, McKeith, Gholkar, & Ferrier, 2004; MacLullich, Deary, Starr, Ferguson, Wardlaw & Seckl, 2004). Instances of chronically high levels of cortisol are classified as illnesses such as Cushing's disease. These stress-level elevations of cortisol associated with Cushing's disease are associated with cognitive decrements (Starkman, Giordani, Berent, Schork, & Schteingart, 2001). While verbal and nonverbal learning, and delayed recall were affected, verbal learning in particular showed a greater negative response. Visual-spatial ability also suffered for learning tasks, though not for recall tasks. These results indicate that chronically high levels of cortisol can have harmful effects on processing in the neocortex and hippocampus. A decrease in cortisol can reverse hippocampal atrophy if Cushing's disease is treated (Starkman, Giordani, Gebarski, Berent, Schork, Schteingart, 1999). However, there can still be lasting impairment even for those with long-term cures (Tiemensma, Kokshoorn, Biermasz, Keijser, Wassenaar, Middelkoop, Pereira, & Romijn, 2013). In particular, cognitive function, reflective memory, and executive functions showed lasting impairment.

Elevated cortisol levels can also affect working memory (Duncko, Johnson, Merikangas, & Grillon, 2009; Elzinga & Roelofs, 2005; Lupien, Gillin & Hauger, 1999), a key component to metacognitive function. Mattarella-Micke, Mateo, Kozak, Foster, & Beilock (2011) has considered cortisol as a measure in studies concerning math anxiety.

In one of their studies, participants solved math problems (modular arithmetic) and both math anxiety and salivary cortisol concentrations were measured before and after the math test. They also completed a working memory capacity measure. Analyses showed that there exists a relation between physiological response and math performance and that this relation is dependent on math anxiety. Participants with high math anxiety showed lower math performance with instances of higher cortisol concentrations, while participants with low math anxiety showed an increase in math performance with higher cortisol concentrations. They further showed that this effect was moderated by individual differences in working memory. Problems that posed low demand on working memory were not affected by their found interaction between math anxiety and cortisol. Participants with low working memory do not extensively use working memory to solve problems, and so their performance was not markedly different on low-demand and high-demand problems. But those with high working memory were impacted by the interaction between math anxiety and cortisol, showing a difference in performance between low-demand and high-demand problems. Furthermore, those with high anxiety as well as high working memory showed lower performance associated with increases in cortisol concentrations. From these findings, Mattarella-Micke et al. suggest that people can achieve higher performance when faced with a more challenging environment, provided they have sufficient cognitive resources and they see their physiological stress response as a motivational cue. Overall, they conclude that the impact of anxiety on working memory causes a related impact on performance.

Conclusion

Math anxiety can negatively impact cognition in a variety of ways. These can be seen through the impact on working memory, and the resulting impact on performance. Negative effects of math anxiety, as a form of psychological stress, can also be seen through the measurement of cortisol. Cortisol provides a more accurate (and honest) measure of psychological stress induced by math tasks than do self-reported perceived stress tests. High levels of cortisol are associated with decreased cognitive function, particularly working memory, and this might indicate that cortisol could serve as a measure of metacognitive function as well.

Chapter 4: Metacognition and Confidence: Comparing Math to Other Academic Subjects (Studies 1 and 2)

Introduction

Metacognition, a form of higher-order thinking, plays an important role in cognitive processing. It impacts several areas within cognitive science, such as attention, memory, perception, comprehension, and problem solving (Kitchener, 1983; Metcalfe & Shimamura, 1994). Metacognition aids intellectual endeavors requiring complex thought processes (Schoenfeld, 1983) and also affects social behavior (Jaccard, Dodge & Guilamo-Ramos, 2005) and decision making (Cohen, Freeman, & Thompson, 1998).

Two components of metacognition are of particular importance in education: the ability to monitor what you know, which acts as a basis for predicting retention, and the control processes that are used to enact study activities (Nelson & Dunlosky, 1991). Students need to use metacognitive control in gauging what they know and deciding what study methods to use (Thiede, Anderson, & Therriault, 2003; Metcalfe, 2009). This process is constantly changing, as students adapt their behaviors in monitoring a learning goal. Self-regulation is necessary for this process (Kornell & Bjork, 2007), thus students must select from a variety of strategies, enacting these strategies in goal-directed activities, and monitoring their progress in using these strategies.

Success of metacognition affects students' academic performance (Hattie, Biggs, & Purdie, 1996; Paris & Paris, 2001; Coutinho, 2008), as well as their ability to communicate what they know about a particular problem. Being able to communicate their level of understanding to instructors is crucial to the learning process. It guides how classroom and self-study materials are constructed, and can affect what strategies students are taught for study and examination.

Metacognition has been shown to play a crucial role in gauging one's own knowledge (Sperling, Howard, Stanley & DuBois, 2004; Schunk & Ertmer, 2000), including specific academic domains such as reading comprehension (Pressley, 2002), math (Pugalee, 2001), science (Schraw, Crippen, & Hartley, 2006), and writing (Pugalee, 2001). Any improvements in metacognition would allow students to better judge what they know and how well they will be able to recall information. This holds much promise for improving student academic performance.

Despite the importance of metacognition, people commonly display glaring overconfidence in their self-perception of their own knowledge and various abilities (Kruger & Dunning, 1999; Dunning, Johnson, Ehrlinger, & Kruger, 2003). Furthermore, people with lower abilities show an even more exaggerated overconfidence. Students in particular often self-report confidence judgments that are unrelated to their actual performance on assessments (Schraw, 1996). Compounding this is students' inability to allocate study times effectively. Methods of self-guided study often result in non-optimal allocation of study time (Son & Sethi, 2009). Improved methods are available, but students generally do not employ them, even though it has been shown that it is possible to use metacognitive control. There is potential for optimal study (Son & Sethi, 2006), but students instead use uninformed metacognitive decisions to structure their study time.

A possible exception to the overconfidence phenomenon is the occurrence of math anxiety (Meece, Wigfield, & Eccles, 1990; Furner & Berman, 2003). Math anxiety (or math phobia) is a fear of math that leads to math avoidance or lower math

performance (Ashcraft, 2002; Ashcraft & Krause, 2007) and has been observed in children and adults alike (Wigfield & Meece, 1988). This sometimes extreme anxiety is harmful in educational and workplace settings (Meece et al., 1990; Furner & Berman, 2003), undermining national and worldwide priorities to emphasize science, technology, engineering, and math (STEM) achievement. Indeed, a recent national report predicts increased demand for STEM professionals in the US as well as an inadequate supply of prepared graduates (STEMconnector, 2013). Performing math tasks in stressful situations, such as during tests, only compounds math anxiety (Beilock, 2008). Math-phobic attitudes of teachers can also be detrimental to students' math achievement, particularly for female teachers and students (Beilock, 2010).

This fear of math implies that there should be a corresponding underconfidence in self-evaluation of mathematical ability. The consistent and persisting documentation of widespread math phobia contradicts the finding that people are generally overconfident. How then can we resolve this paradox? We wish to determine if students are as overconfident in math as they are in other academic domains, or if is math an exception to an otherwise global overconfidence.

Past findings indicate that females generally lag behind their male counterparts on standardized test performance in math (Brown & Josephs, 1999). This is particularly true among high school and college students (Hyde et al., 2006). This gap does appear to have narrowed in recent years (Else-Quest, Hyde, Shibley, Marcia, 2010). However, attitudes toward math between genders still follow differing patterns, and females are more likely to feel intimidated by math than are males (Jakobsson, 2012; Brown & Josephs, 1999). This lack of confidence often leads to a self-fulfilling lag in performance (Brown & Josephs, 1999; Kiefer & Sekaquaptewa, 2007) that can lead to gaps in performance between genders.

We wish to explore is if overconfidence generalizes to all domains of academic knowledge and ability, or if it is domain specific. If there exist confidence differences among various academic subjects, this suggests that overconfidence is domain specific and not a general phenomenon that is implied by the findings of Kruger & Dunning (1999) and Dunning et al. (2003). If overconfidence is a global phenomenon, we would expect to see overconfidence in students' ratings across various academic domains. If metacognition is instead domain specific, we would then expect to find differences in overconfidence among academic domains. In the light of math phobia, we would expect to see underconfidence rather than overconfidence in math tasks, in contrast to other domains.

We also seek to determine if metacognitive ability differs over gender as well, keeping in mind that female students show greater math phobia than males. Finally, we compared metacognitive judgments before and after an intervention, namely taking a test, to determine if students are able to improve their metacognitive judgments. We expected to see improvements, as people could re-evaluate their metacognitive estimates after being exposed to more information in the intervention. This would be consistent with Bayesian accounts of cognition, in that people would be updating their hypothesis of ability based on new observations (Jones & Love, 2011; Heit & Erickson, 2011).

Study 1A

We considered test performance, confidence, and calibration in predicted scores. Three comparisons will be highlighted. The first is the comparison among the three different SAT II Subject Tests to assess if overconfidence is a domain specific or general phenomenon. While predictions (estimates before taking an assessment) provide a measure of general confidence within a subject, postdictions (estimates after taking an assessment) provide a more accurate and comparable measure of metacognitive ability to evaluate knowledge. The use of SAT II Subject Test sample questions gave participants a reference for difficulty level of the assessment before they take it. However, it might have been some years since the participants have taken these, and some participants may have chosen to take a different selection of subject tests than the ones presented in this experiment. Use of retrieval fluency and recognition heuristics would negatively affect metacognition, both for past experience and future performance (Benjamin, Bjork, & Hirshman, 1998). The use of postdictions brings all participants to a more equitable level of familiarity with the test material before making a judgment of ability. As such, calibration was determined by comparing postdicted estimates of performance with actual scores of performance. If overconfidence is domain specific, we then expect that metacognitive performance would differ among different domains, and that there would be a higher rate of underconfidence within math. If metacognition is domain general, then a similar level of overconfidence should be observed across all three assessments.

The second comparison will be one made between genders. Males were expected to show higher confidence ratings in math than females. Third is the comparison between predictions and postdictions for performance on a task. This allows us to determine if participants improved their metacognitive judgments after completing a task. We expect that postdictions for performance on a task will be more accurately calibrated than predictions for the same task, and results reflected this.

Study 1A Method

Participants There were 31 participants in this experiment: 17 female and 14 male. All were UC Merced undergraduates (mean age = 19.03, SD = 0.98) who took the experiment as a form of extra credit in one of their introductory Psychology or Cognitive Science classes.

Tasks and Materials Participants took three tests: a biology, literature, and math test, each consisting of 15 questions. Participants were told before the experiment that they would be taking tests based on SAT II Subject Tests content. Before each test, participants were asked to provide a predicted score (out of 15) for how well they would do. After taking each test, participants provided a postdicted score for how well they thought they performed. They were not told their actual scores on tests.

Study 1A Results

Key descriptive results for all participants are shown in Figure 1. The leftmost bar for each category represents average predicted score, the middle bar represents average actual test score, and the rightmost bar represents average postdicted score. Average performance across all tests was 40%. Participants showed general

overconfidence in predicted scores before each test. Overconfidence generally persisted in postdicted scores, although drastic reductions in residual magnitudes show that participants were better able to assess their ability after each test, providing evidence for improvement in metacognitive judgment of ability ($t = 3.30$, $df = 60$, $p < 0.0001$). The only test in which participants showed slight underconfidence was biology.

Notably, participants showed high overconfidence in math. The residual for average predicted math score was 35%. This was higher than the residuals for both biology (2.6%) and literature (19%).

Results by gender are shown in Figures 2 and 3. Notable was the difference in calibration between genders. Overall, females were more accurate in self-estimates of ability. Differences between their predictions and scores averaged 11%, compared to 29% for males. Similarly for postdictions, females misestimated their performance by an average of 5% while males misestimated by an average of 14%. Males were generally overconfident both before and after taking each assessment. Overall, females had lower measures of overconfidence. Within literature and math, females began with overestimates of their ability, but their postdictions were more calibrated. Within biology, females actually started underconfident and became even more so after taking this assessment.

Both genders show little trace of math phobia, as shown by their predominant overestimates of performance. Average prediction and postdiction residuals in math were 27% (overconfident) and –5% (underconfident) for females and were 42% and 15% for males. Though participants were generally overconfident with their predictions, they were able to improve their metacognitive judgment accuracy significantly in this domain. Males showed the most marked improvement in calibration in math, and females actually changed their judgments from being overconfident to predominantly underconfident.

In a three-way, predicted versus actual score × academic subject (biology or literature or math) × gender (male or female) ANOVA, there was a main effect of gender $F(1,29) = 4.48$, $MSE = 13.69$, $\eta^2 = 0.08$, $p < 0.05$. There was also a significant main effect of predicted (*mean* = 59.00) versus actual (*mean* = 40.22) score, $F(1,29) = 36.61$, $MSE = 10.1$, $\eta^2 = 0.50$, $p < 0.0001$, indicating overconfidence in predictions. There was also a significant main effect of academic subject, $F(2, 116) = 6.59$, $MSE = 5.33$, $\eta^2 = 0.08$, $p < 0.01$. Notice that scores were lowest overall in math. There was a significant interaction between these two variables, $F(2, 116) = 16.80$, $MSE = 5.33$, $\eta^2 = 0.20$, $p < 0.0001$, indicating that degree of overconfidence depended on academic subject. Overconfidence was greatest in math (*predicted score* = 62.15, *actual score* = 27.53). We are careful not to over-interpret the interaction, as actual scores also varied by academic subject. There was also a significant interaction between gender and predicted versus actual score, $F(1,29) = 8.19$, $MSE = 10.1$, $\eta^2 = 0.11$, $p < 0.01$, providing further evidence that overconfidence depended on gender. The remaining main effects and interactions were not significant.

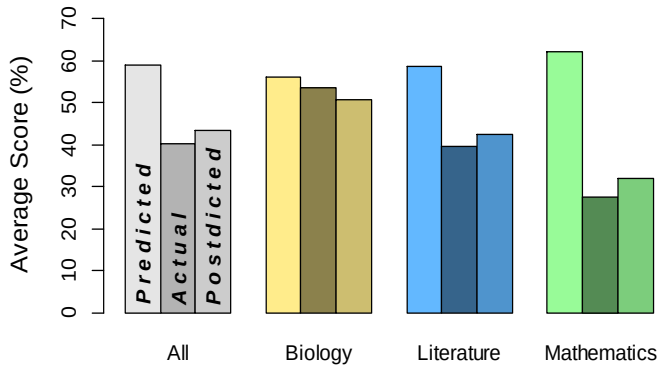

Figure 1: Overall results by test.

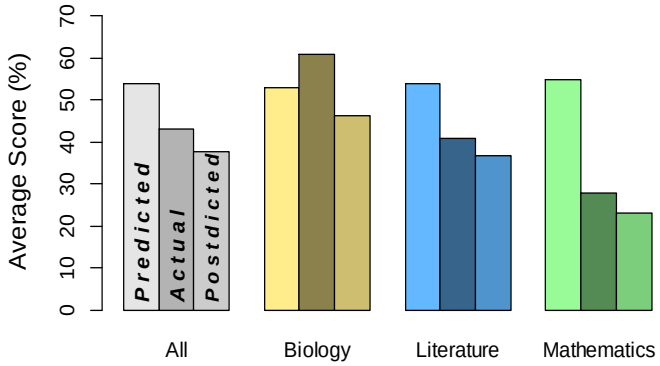

Figure 2: Overall results for females.

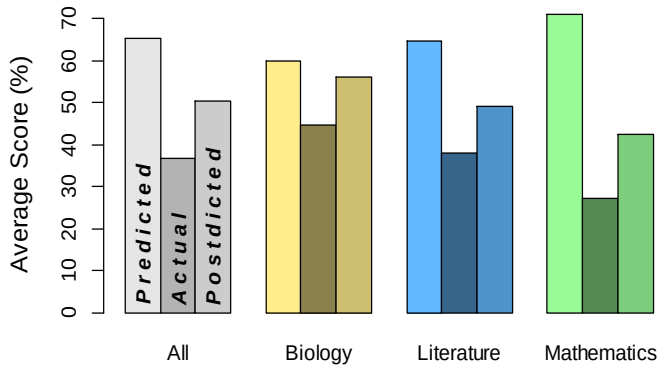

Figure 3: Overall results for males.

We also conducted a similar analysis on postdicted scores (*mean* = 43.44) and actual scores. This ANOVA revealed a main effect of academic subject, $F(2,116) = 36.28$, $MSE = 4.85$, $\eta^2 = 0.37$, $p < 0.0001$, as well as a main effect of gender $F(1, 29) = 4.48$, $MSE = 15.49$, $\eta^2 = 0.09$, $p < 0.05$. There was also a significant interaction between gender and postdicted versus actual scores, $F(1, 106.27) = 17.55$, $MSE = 6.05$, $\eta^2 = 0.37$,

$p < 0.001$, again showing gender differences in overconfidence. Remaining main effects and interactions were not significant.

Note that although scores were lowest for math, this assessment was not designed to be more difficult than the other subject tests. In fact, it had the lowest difficulty level. During pilot experiments, test questions were chosen using difficulty ratings provided by College Board. Although we originally chose a variety of easy, medium, and difficult questions for each subject test, performance on this balanced math test was so poor that we substituted easier questions in place of all medium and difficult questions. Thus, the severe overconfidence observed in math is not a result of higher test difficulty level compared to other academic subjects.

Figures 4 and 5 show calibration slopes by domain. The dashed line represents the equation $y = x$ (predicted score = actual score) is used to convey perfect calibration. The closer a line is to this dashed line, the better the calibration. For the predicted scores, there are apparent subject differences, e.g., the slope is highest for math, indicating the highest level of sensitivity to actual performance, and the slope is actually slightly negative for biology. Each domain slope more closely follows the calibration line $y = x$ for postdicted scores, showing that that participants were better able to judge their ability after taking each assessment. Each of models also crosses $y = x$, switching from overconfidence to underconfidence as test performance increases. In addition, correlations between actual scores and residuals calculated from estimated scores ($r = -0.63$ for both predicted and postdicted residuals) reveal that higher scores are associated with lower residuals. These findings support the previous work by Kruger & Dunning (1999), Dunning et al. (2003) and Schraw (1996) and show that people with low test scores generally exhibit overconfidence, while people with high test scores are better able to judge their ability. Thus higher performing students tend to be better judges of their ability than are lower performers.

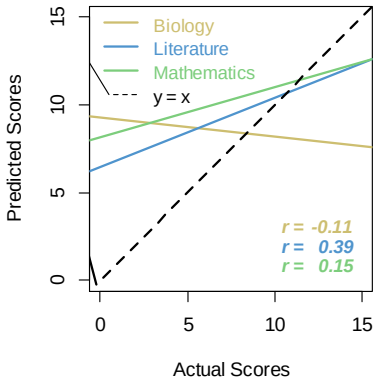

Figure 4: Overall results for predictions.

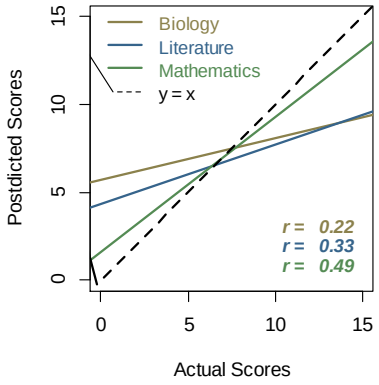

Figure 5: Overall results for postdictions.

Gender differences were most striking within math. Figure 6 and 7 show calibration models by gender for both predictions and postdictions. Males made predictions with almost no calibration ($r = 0.02$), and females were overconfident overall with the predictions. Despite this, both genders were able to make much more accurate

postdictions. In fact, these postdiction models were the best of any of the observed estimates of ability in this experiment when compared to other subjects.

Study 1A Discussion

The results of this study support past findings that people are generally overconfident in their abilities, although overconfidence does not appear to be exactly the same across domains. This was shown in the varying judgments of ability across academic domains. While both males and females are generally overconfident, females tend to be better calibrated in judging their domain knowledge. Postdictions were significantly lower than predictions, showing that people are able to recalibrate their metacognitive judgments towards more accurate judgments after attempting an assessment.

In addressing the paradox of general overconfidence alongside the seeming exception of math phobia, we saw that overconfidence was particularly high in predicted scores for math assessment. This led us to question whether math phobia was present.

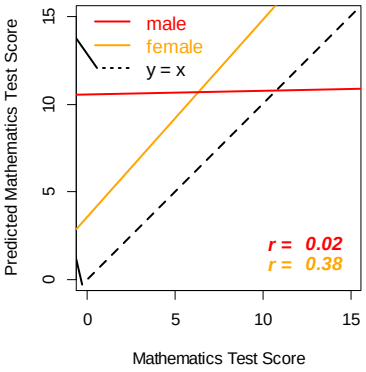

Figure 6: Math results by gender (predictions)

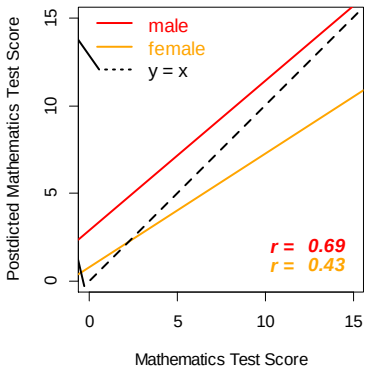

Figure 7: Math results by gender (postdictions).

Although there was a higher incidence of overconfidence in mathematics, participants showed the greatest beneficial adjustment of metacognitive judgment miscalibration for mathematical ability. All participants were successfully able to recalibrate their estimates towards more accurate judgments of domain knowledge after an assessment.

We have replicated this severe overconfidence in math in other experiments, although gender no longer reached the level of statistical significance. Thus high math overconfidence is not specific to college-level students: In a subsequent experiment (n = 40), this result was replicated at a local high school using the same experimental design. In another experiment with college students (n = 46), we extended our findings by using the same experiment presented here, but also including Likert scale measures of confidence for each domain, as well as an adapted math Anxiety Rating Scale (MARS) survey (Alexander & Martray, 1989). Initial results indicate that there does exist math phobia, as we observed MARS ratings similar to other college populations identified as math anxious. We expected higher anxiety ratings to be linked with underestimates of ability. We did not observe this. Instead, math phobia moderated overestimates of ability to be less extreme, although overconfidence still persisted. This is a possible

explanation for the coexistence of math phobia and overconfidence in mathematical ability. Further plans include the replication of studies within actual classroom settings in which participants must judge their ability on class assessments.

Does a higher confidence in one subject over another really indicate domain specificity rather than generality? If so, this suggests we may be using different metacognitive methods for different domains such as sciences versus the humanities. Alternatively, we might be using one overarching metacognitive ability that uses different cues and leads to different results across domains.

Two components of metacognition are particularly relevant for successful learning: self-monitoring, e.g., assessing performance, and self-regulation, e.g., choosing what and how to study (Metcalfe, 2009; Nelson & Dunlosky, 1991; Thiede, Anderson, & Therriault, 2003). Metacognition has been found to be crucial for calibration of self-knowledge of ability (Schunk & Ertmer, 2000; Sperling, Howard, Stanley & DuBois, 2004), in domains and tasks such as mathematics (Pugalee, 2001), science (Schraw, Crippen, & Hartley, 2006), reading (Pressley, 2002), and writing (Pugalee, 2001). If people are unable to assess their performance accurately, then it is unlikely that they will be able to learn optimally (see Townsend & Heit, 2011, for a related argument). Any improvements in metacognition would allow learners to better judge what they know and how well they will be able to learn information and recall it later.

Our focus in the present research is self-monitoring in math. In addressing the topic of metacognition and math, it is important to consider whether metacognition is domain-general or domain-specific. In other words, to what extent are there general points to be made about people's metacognitive abilities, potential for error, and underlying mechanisms across subject domains, and to what extent are there distinctive points to be made for particular domains? Do the difficulties that learners face with math reflect general issues with metacognition, or something special about math?

Though there is ongoing debate on whether metacognition is domain-specific or domain-general, there is no debate on the importance of metacognition in any learning process. In this paper we will focus primarily on possible differences between metacognition in mathematics learning when compared to other domains. In exploring this we find it relevant to also address math anxiety, which can affect both domain performance and metacognitive performance.

Based on past research, it remains unclear whether metacognitive performance is similar across domains. Metacognition research typically addresses a single subject domain in isolation, and those few studies comparing metacognition across domains have shown mixed results (e.g. Veenman, Van Hout-Wolters, & Afflerbach, 2006). Though these past studies have used varying methodologies and assessments for measuring metacognition, results still bear on this issue of domain generality versus specificity and provide two views which we must consider.

Domain-general views of metacognition

Some studies support domain-generality of metacognition, treating it as a skill that can be applied across different content areas (e.g. Schraw, 1996; Halpern, 1998; Veenman & Verheij, 2001). In these cases, domain-general metacognitive skills are distinguished from domain-specific knowledge. This framework assumes that cognitive

skills can be domain-specific whereas metacognitive skills can be applied across even unrelated domains. Interestingly, metacognitive ability appears unrelated to IQ (Alexander, Carr, & Schwanenflugel, 1995). Rather, metacognitive skills are assumed to improve along with domain knowledge.

Metacognitive skills are further related to domain knowledge in that metacognitive skill can aid learners those with low ability or knowledge. For instance, Swanson (1990) showed that metacognitive ability compensated for IQ in a comparison between fifth and sixth grade student problem solving ability. Ability to solve problems was unrelated to IQ, while those with higher metacognitive ability were better able to solve problems than those with lower metacognitive ability. This result suggests that metacognition can be applied flexibly across tasks and thus is a domain-general skill.

This domain-general view of metacognition is consistent with the unskilled and unaware phenomenon, in which people show domain-general overconfidence in their abilities, with low performers showing greater overconfidence than higher performers (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999). This phenomenon has been shown for students predicting performance on laboratory tests ranging from logical reasoning to grammatical knowledge and sense of humor. Studies with academic content in classroom settings also exhibit this phenomenon (Maki & Berry, 1984; Miller & Geraci, 2011). Miller and Geraci also demonstrated that the lowest performers were overconfident in exam score predictions, but they were also less confident in these predictions than were the highest performers. Thus, although the unskilled might be more aware than once thought, they still demonstrate overconfidence nonetheless. Furthermore, while people with high performance might demonstrate slight underconfidence, people with lower performance have even more exaggerated overconfidence. From this, one might predict that students who are struggling in math are particularly overconfident.

Domain-specific views of metacognition

In contrast, some investigations of individual differences in metacognition point towards domain specificity. For example, Kelemen, Frost, and Weaver (2000) found that metamemory accuracy was task specific for university students. They tested memory monitoring performance across four metacognitive tasks: ease of learning judgments for Swahili-English word pairs, feeling of knowing judgments for general knowledge questions, judgments of learning for unrelated English word pairs, and text comprehension monitoring for narrative texts. While they found individual differences in memory and confidence that remained constant across tasks, individual differences in metacognitive accuracy changed for each task. Glaser, Schauble, Raghavan, and Zeitz (1992) provided evidence that metacognition can differ based on task. They found that metacognitive strategies of university students varied across discovery learning tasks. In their comparison of several reasoning tasks, they further found variability in metacognitive performance for components of problem solving. There are a variety of alternative but equally successful problem solving strategies and a variety of metacognitive approaches for these problems. In general, successful problem solvers use metacognitive strategies more often than less successful problem solvers, but there is no one set of metacognitive strategies that led to successful problem solving. They also

found that domain content and context led to variation in use of metacognitive strategies, and that particular metacognitive skills were associated with specific learning success or failure within particular domains and contexts.

Still others have suggested that metacognition might be domain-specific early in development, beginning as reflective self-analysis of cognition. For example, Paris and Byrnes (1989) suggest that such self-directed reflection develops in children as self-corrections, and this behavior becomes more prominent as children get older. As children develop self-regulation for individual tasks, then gradually learn to apply general self-correction skills across a variety of tasks. Similarly, Landsmann and Karmiloff-Smith (1992) suggested that this reflection results in the restructuring of self-knowledge that increases theoretical understanding of one's cognition. This restructuring starts to differentiate and separate various domains of knowledge. Both these views support a theory of first domain-specificity of metacognition that eventually extends to be a domain-general skill. Then as metacognitive skill improves, this skill can be applied across a variety of domains. Veenman and Spaans (2005) provided evidence for this in their findings of first domain-specificity in metacognitive skill in the first year of high school then domain-generality later for third year high school students. First-year and third-year students solved math problems while thinking aloud and also performed an inductive learning biology task. Metacognitive skillfulness was measured based on enactment of metacognitive behaviors (e.g. entirely reading a problem statement, selection of relevant information needed to solve the problem, monitoring the on-going problem-solving process, checking the answer, reflecting on the answer). A difference in metacognitive skillfulness was observed when comparing the two groups of students in that metacognitive skills are at first domain-specific for first-year students, while they are domain-general for third-year students.

Previous research shows connections between metacognition and math anxiety. Math anxiety (or math phobia) is a fear of math that leads to math avoidance or lower math performance (Ashcraft, 2002; Ashcraft & Krause, 2007). This sometimes extreme anxiety is harmful in both educational and workplace settings (Furner & Berman, 2003; Meece, Wigfield, & Eccles, 1990). Performing math tasks in stressful situations, such as during tests, only compounds math anxiety (Beilock, 2008; Beilock & Carr, 2005). This anxiety can start early in children's education, with elementary school students already showing harmful effects of math anxiety on their math achievement (Ramirez, Gunderson, Levine, & Beilock, 2013). Math anxiety interrupts cognitive processing through its interference with working memory, and this is what can cause people to show lower performance under pressure (Beilock & Carr, 2005). While math anxiety does not appear to affect simple math tasks such as single digit addition, it does affect decision-making processes for number sense and any task that required procedural aspects of arithmetic (Dehaene, 1997; Ashcraft, 2002). The tasks requiring use of working memory are adversely affected by math anxiety. Students who are highly math anxious tend to make more errors in timed problems than did those with low math anxiety. This is also consistent with Eysenck and Calvo's (1992) model of general anxiety effects, in which general anxiety disrupts working memory through preoccupation with thoughts and attention given to worry instead of to the current task. This preoccupation is a second

task that places a heavier load on working memory, a component of cognition that is used in metacognition (Shimamura, 2000).

The widespread evidence for math anxiety suggests the view that math may be uniquely problematic compared to other academic subjects. There does not seem to be a corresponding body of evidence for, say, literature anxiety or even biology anxiety. Generalizing this point, we would expect that if students fear math, then they should generally have low confidence in math compared to other subjects and a corresponding difference in metacognitive ability. Indeed, Ashcraft (2002) found evidence for this point in terms of strong negative correlations between math anxiety and self-confidence in math.

Overview of experiments

Accounts of domain specificity are consistent with the idea that math is uniquely problematic. Contrasting this view that math is unique are accounts of domain generality, including the unskilled and unaware phenomenon. So are students underconfident in math, as would be expected from the math is unique view, or are students overconfident in math, as would be predicted by the unskilled and unaware view? Note that we present these as opposing views, but the predictions from these views were derived by ourselves.

Our purpose is not to examine whether math-phobic students are less confident than non-math-phobic students or whether students are less confident in math than in other subjects. Rather, we focus on calibration of metacognitive judgments, that is, whether students are under- or over-confident relative to their performance. In addition to general measures of calibration, we also compared calibration across academic subjects. We assess both absolute calibration, in which we simply measure how well subjective scores matched objective scores, as well as relative calibration, assessing whether participants with higher subjective scores had correspondingly higher objective scores. Other than looking at confidence for students with lower versus higher scores, it was not our aim to examine individual differences.

In two studies, we assessed confidence in math as well as other academic subjects (biology, literature). These studies were conducted based on approval by the Institutional Review Board (IRB) for our institution (University of California, Merced). High school (Study 1) and college (Study 2) students took standardized tests and estimated their performance. Following the view that math is unique, we would expect underconfidence in math relative to the other subjects and likely worse metacognitive calibration. In contrast, following the unskilled and unaware view, we would expect similar overconfidence in math compared to other subjects. Past research shows that high school students are generally math anxious (Hembree, 1990; Ma, 1999). From this, we expected that high school students in Study 1 would be math anxious, and we did not directly measure anxiety. However, in Study 2, we included a standard measure of math anxiety. We note again that it was not our purpose to examine individual differences in anxiety but rather to look at groups of students who were most likely math anxious overall, and our findings are limited to these groups.

Study 1B

Method

Participants in this study attended a summer program at a diverse public high school in California. They made two estimates: predictions (before each test) of their performance as well as postdictions (after each test). Multiple choice tests were adapted from teachers' materials used at that grade level in this school, giving students a basis for making predictions, with even more information when making postdictions.

Our main focus was to compare calibration of estimates about math to the other two subjects. We also compared predictions to postdictions, allowing us to determine if metacognitive judgments improved after completing a test, as would be expected from previous research (Dunning et al., 2003; Kruger & Dunning, 1999).

Participants. There were 40 participants (25 female, 15 male). All were students (*mean age* = 15.27, *SD* = 0.55) at a local high school in the Central Valley of California, who took the study for extra credit in their summer school class (Algebra 1). A majority of these students had failed math the previous academic year. The school is located in a rural area; its student population is 72% Hispanic-Latino, 21% White-Caucasian, 3% Asian-Pacific Islander, 1% African American, and 3% of other ethnicity.

Materials and Procedure. Each participant took three computer-based tests (biology, literature, mathematics, in randomized order), each test consisting of 15 questions with 15 minutes allowed per test. Questions were normed so that the overall level of difficulty across tests was comparable, to avoid ceiling and floor effects and to assure variance in scores. Questions left unfinished within the allotted time were scored as incorrect. Participants were told that these tests were similar to those from their current classes. Question style was multiple choice with five answer choices. Before each test, participants provided a predicted score (number of questions correct) for how well they would do. After each test, participants provided a postdicted score for how well they thought they had performed. They were not told their actual scores.

Study 1B Results and Discussion

Key descriptive results are in Figure 1. The leftmost bar for each category represents average predicted score, the middle bar represents average actual test score, and the rightmost bar represents average postdicted score. The general pattern is that predictions are substantially greater than actually performance, and postdictions are somewhat greater than actually performance. This overconfidence is particularly striking for predictions about math performance. We do not show breakdown by gender, however, predicted, actual, and postdicted scores averaged about 10% higher for males.

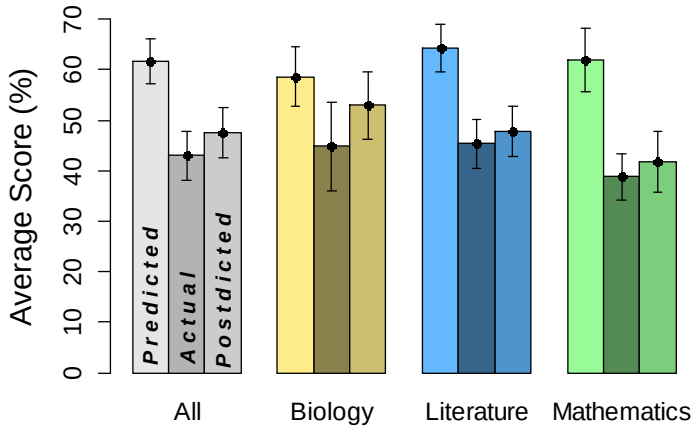

Figure 1. Predicted, actual, and postdicted scores by domain. Leftmost bars represent predicted scores, middle bars domain performance, and rightmost bars postdicted scores, each with standard error.

First, we examined predictions in a three-way, predicted versus actual score × academic subject × gender, ANOVA. There was a significant main effect of predicted (*mean* = 61.7) versus actual (*mean* = 43.0) score, $F(1, 38) = 54.45$, $MSE = 6.00$, $\eta^2 = 0.59$, $p < 0.0001$, indicating overconfidence in predictions. The academic subject variable did not reach statistical significance, $F(2, 152) = 4.14$, $MSE = 4.11$, $\eta^2 = 0.05$. However, there was a significant predicted versus actual score × academic subject interaction, $F(2, 152) = 10.31$, $MSE = 4.11$, $\eta^2 = 0.11$, $p < 0.0001$, implying that overconfidence differed by academic subject. Notably, participants showed the highest degree of overconfidence in mathematics (*predicted score* = 62.0, *actual score* = 38.9). There was also a main effect of gender, $F(1, 38) = 5.88$, $MSE = 17.41$, $\eta^2 = 0.13$, $p < 0.05$, but remaining interaction terms were not statistically significant, $F < 1$. Hence, degree of overconfidence did not depend on gender, although it may be that the sample size did not yield enough power to fully address this point.

We also conducted a comparable analysis on postdicted scores (*mean* = 47.5). In this three-way ANOVA, there was a significant effect of academic subject, $F(2, 152) = 21.00$, $MSE = 4.06$, $\eta^2 = 0.21$, $p < 0.0001$. Note that biology had the highest values overall and mathematics had the lowest. There was also a significant main effect of gender, $F(1, 38) = 5.10$, $MSE = 19.74$, $\eta^2 = 0.12$, $p < 0.05$. The remaining main effect and interaction terms were not statistically significant, $F < 1$. Hence, we did not see significant overall overconfidence on postdictions, and overconfidence did not depend on academic subject or gender.

Table 1
Correlations between participant-produced estimates and performance by domain.

	Biology	Literature	Math
Prediction versus actual score	0.32*	0.10	0.56***
Postdiction versus actual score	0.39*	0.21	0.64***

Note: *$p < 0.01$; * $p < 0.05$; ***$p < 0.001$ (as compared to 0)

The preceding analyses focused on absolute calibration, namely how well subjective scores matched objective scores, on average. We next examined relative calibration, namely whether participants with higher subjective scores had correspondingly higher objective scores, measuring relative calibration in terms of correlation coefficient, *r*, across all participants (Table 1). Relative calibration is particularly strong for math, and particularly weak for literature, with biology falling between.

We also tested differences in correlations both across domain and within domain. Steiger *z*-tests of independent correlations showed that the prediction versus actual score correlations are significantly different between math and literature ($z = 2.28$, $p < 0.02$). Similarly, postdiction versus actual score correlations are significantly different between math and literature ($z = 2.36$, $p < 0.02$). Williams *t*-tests of dependent correlations reveal that differences between prediction versus actual score and postdiction versus actual score are significant for all of biology ($t = –3.19$, $p < 0.0029$), literature ($t = –3.13$, $p < 0.0034$), and math ($t = –2.28$, $p < 0.028$). Thus calibration significantly improved in postdictions on all tests when compared to calibration of predictions.

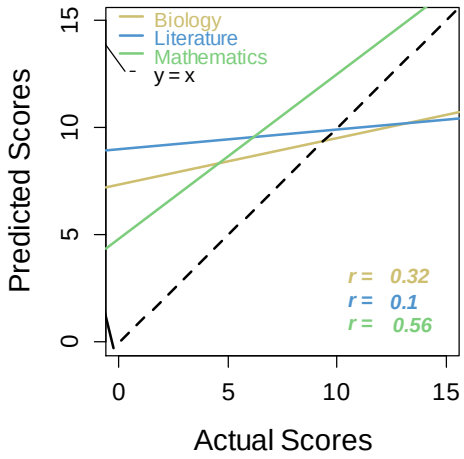

Figure 2. Regression lines for actual versus predicted scores by academic subject. Regression lines that more closely follow the main dashed line (perfect calibration) indicate better metacognitive calibration for that academic subject.

Linear regression calibration curves also illustrate this point (Figure 2). Regression lines that more closely follow the main dashed line (perfect calibration) indicate better metacognitive calibration for that academic subject. Regression lines for

biology and literature are shallow, showing little sensitivity to actual performance. These lines show the usual unskilled and unaware pattern of overconfidence at the lowest level of performance and underconfidence at the highest level (Dunning et al., 2003; Kruger & Dunning, 1999). Calibration for math in this case seems to take unskilled and unaware one step further, in that even the high performers over-estimated their ability.

In general, results more closely supported the unskilled and unaware view rather than the math is unique view. Students were generally overconfident in all three academic subjects, at least on predictions if not postdictions. The only evidence we found suggesting a difference for math is that relative calibration for math was actually the best and overconfidence was the greatest. Overall performance was somewhat lower for female students than for males, but so were predictions and postdictions, so their overconfidence was no different.

Study 2

We turn to another study, attempting to replicate and extend key findings from Studies 1A and 1B, which might have been due to idiosyncrasies of the particular student sample or test instruments used. In Study 2, we conducted a similar study on college students, using sample SAT test questions. We would have expected high school students in Study 1 to be math anxious, because most high school students show some math anxiety (Hembree, 1990; Ma, 1999; Maloney & Beilock, 2012), and most of the students in our study had previously failed math classes. In Study 2, we included a standard measure of math anxiety adapted from a shortened Math Anxiety Ratings Scale (MARS, Alexander & Martray, 1989). Although we did not directly measure math anxiety for students in Study 1, we replicated this study within the same school population the following summer, and the average shortened MARS score was 77. This study also included additional measures but otherwise replicated findings from the high school study presented here. For comparison, Ashcraft and Moore (2009) found an average shortened MARS score of 61 across several college samples identified as math anxious.

Study 2 Method

Participants. There were 46 participants (28 female, 18 male) in this study. All were UC Merced undergraduates (*mean age = 19.96*, *SD = 1.75*) who received extra credit in their introductory psychology or cognitive science classes for their participation. The UC Merced undergraduate population is 40% Hispanic-Latino, 29% Asian-Pacific Islander, 17% White-Caucasian, 7% African American, and 7% other ethnicity.

Materials and Procedure. Participants took three tests (biology, literature, mathematics), each with 15 questions. Again, questions left unfinished within the allotted time were scored as incorrect. Participants were told that these were based on SAT II Subject Tests (which most students have taken). Students completed three assessments derived from past questions released by the College Board, making predictions and postdictions as in Study 1. Test questions were normed to avoid ceiling and floor effects. Then, participants answered 23 questions about math anxiety from a variant of the shortened MARS (Alexander & Martray, 1989), using a rating scale ranging from 1 – "no anxiety" to 5 – "very high anxiety", with possible scores from 23 to

115. (We dropped two of three questions from the standardized shortened MARS with nearly identical wording.)

In pilot studies, in an effort to design tests of equal difficulty, items of comparable difficulty for each test were selected using individual question ratings (easy, medium, difficult) provided by College Board. However, performance floor effects on the pilot math tests were so pronounced that easier questions were substituted in an effort to bring average performance closer to the level as the literature and biology tests.

Study 2 Results and Discussion

The results largely replicated Study 1 (see Figure 3), most notably in terms of general over-confidence of predictions compared to actual scores, most notably for math.

In a three-way, predicted versus actual score × academic subject × gender ANOVA, there was a significant main effect of predicted (*mean* = 51.3) versus actual (*mean* = 40.4) score, $F(1,44) = 19.70$, $MSE = 9.31$, $\eta^2 = 0.31$, $p < 0.0001$, indicating overconfidence in predictions.

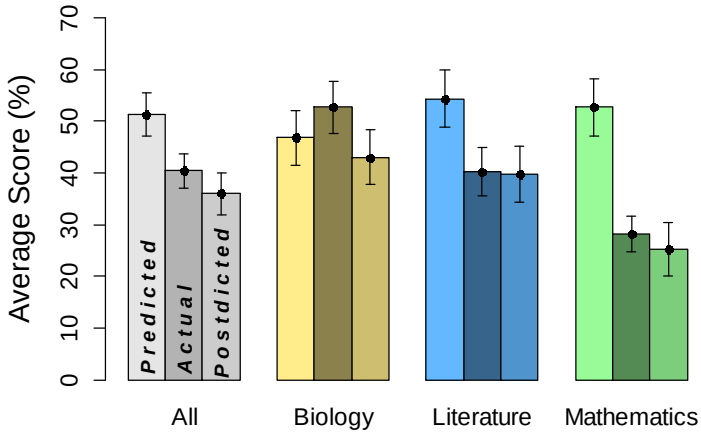

Figure 3. Predicted, actual, and postdicted scores by domain. Leftmost bars represent predicted scores, middle bars domain performance, and rightmost bars postdicted scores, each with standard error.

There was also a significant main effect of academic subject, $F(2, 176) = 7.19$, $MSE = 4.18$, $\eta^2 = 0.09$, $p < 0.0001$; scores were lowest overall in math. There was a significant interaction between these two variables, $F(2, 176) = 29.58$, $MSE = 4.18$, $\eta^2 = 0.22$, $p < 0.0001$, indicating that degree of overconfidence depended on academic subject. Overconfidence was greatest in mathematics (*predicted score* = 52.8, *actual score* = 28.3), however we are careful not to over-interpret the interaction because actual scores also differed by academic subject. The remaining main effect (gender) and interaction terms were not statistically significant, $F < 1$. Hence, the finding of overconfidence, particularly in math, did not depend on gender.

We also conducted a comparable analysis on postdicted scores (*mean* = 36.0) and actual scores. This ANOVA revealed a main effect of postdicted versus actual score, $F(1,44) = 4.48$, $MSE = 6.85$, $\eta^2 = 0.09$, $p < 0.05$, indicating that participants were

slightly yet significantly underconfident overall. With that said, given the overall drop from predictions to postdictions, we see this as reflecting that participants were better calibrated after taking the test, as in Study 1. (A third ANOVA comparing predicted and postdicted scores showed a main effect of predicted versus postdicted.) There was a main effect of academic subject, $F(2, 176) = 37.52$, $MSE = 3.70$, $\eta^2 = 0.41$, $p < 0.0001$. There was also a significant interaction between academic subject and postdicted versus actual score, $F(2, 176) = 3.12$, $MSE = 3.70$, $\eta^2 = 0.02$, $p < 0.05$, indicating a difference in overconfidence by academic domain. Though all medium and difficult questions were removed and replaced with easy ones (based on College Board question ratings), performance remained lower for the math test compared to the other two subjects. The remaining main effect (gender) and interaction terms were not statistically significant, $F < 4$. Even after replacing all questions with those rated as easy by the College Board, and thus creating a minimally difficult assessment that incorporated actual SAT questions, math test scores were consistently lower than biology or literature test scores. Analyses were also conducted using a sample of these math questions by selecting seven questions that participants scored best on. However, this did not change our findings. Again as in Study 1, the finding of overconfidence, particularly in math, did not depend on gender.

Table 2

Correlations between participant-produced estimates and performance by domain.

	Biology	Literature	Math
Prediction versus actual score	0.01	0.12	0.47***
Postdiction versus actual score	0.41**	0.38**	0.35*

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (as compared to 0)

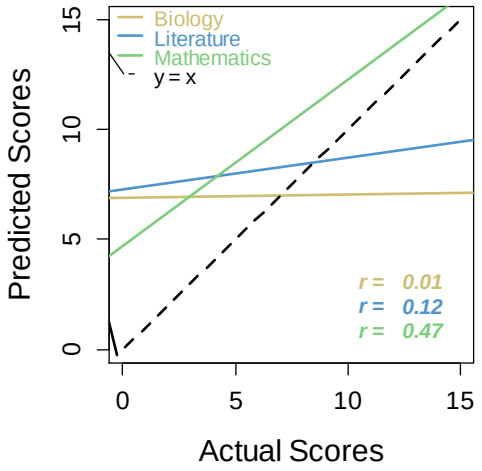

Figure 4. Regression lines for actual versus predicted scores by academic subject. Regression lines that more closely follow the main dashed line (perfect calibration) indicate better metacognitive calibration for that academic subject.

Analyses of relative calibration (Table 2, Figure 4) yielded results similar to Study 1. Correlations between predicted and actual scores were highest for math, although, based on a Steiger *z*-test of independent correlations, there was only a significant difference in correlations between predicted and actual score for biology and predicted versus actual score for math ($z = 2.32$, $p < 0.02$). Correlations for postdicted scores were approximately the same for all three academic subjects, and there were no significant differences among these correlations. Williams *t*-tests of dependent correlations reveal that there were significant differences in predicted versus actual and postdicted versus actual correlations for both literature ($t = –3.21$, $p < 0.0025$) and biology ($t = –3.82$, $p < 0.00042$). Thus calibration was significantly different before and after the test for both literature and biology. However, this pair of correlations for math showed no significant difference, hence there was a lack of evidence for improved calibration in math. Again, the regression lines for biology and literature showed overconfidence at the lowest level of performance and underconfidence at the highest level, revealing again the general unskilled and unaware pattern. In contrast, students were simply overconfident in general for math.

Students were clearly math anxious overall, with an average adapted shortened MARS score of 73/115 ($sd = 16.8$). Comparatively, Ashcraft & Moore (2009) found an average MARS score of 61/125 using a test instrument with two more questions, across several college samples identified as math anxious. Further analyses suggested that more anxious participants had lower performance and lower levels of overconfidence; however, inferential tests did not reach statistical significance. Therefore we simply conclude that students generally experienced math anxiety. We measured a Cronbach's alpha of 0.94 for the version of the shortened MARS math anxiety that we used, indicating very good internal consistency. This is comparable to the Cronbach alphas Plake and Parker (1982) found for the original full-length 98-question MARS instrument (alpha = 0.97) as well as the shortened 24-question version MARS-R (alpha = 0.98).

Overall, the results were consistent with the unskilled and unaware view, in that there was global overconfidence in predictions across subject domains. The finding that even the students who performed best in math were overconfident (see Figure 4) is not consistent with unskilled and unaware, however. Instead of observing underconfidence in the best performers, as has typically been displayed by the unskilled and aware phenomenon, we see persisting overconfidence even in the highest performing participants. In terms of math being unique, it was unique here in the sense that overconfidence was the greatest and metacognitive calibration was the best. Neither of these results would be predicted from the idea that anxiety is particularly high for math.

General Discussion

Our basic conclusion is that students aligned in some ways to both the unskilled and unaware phenomenon and the idea that math is unique. In both studies, students over-estimated their performance in their predicted scores for all domains, though their calibration did improve for postdictions. This provides support for the domain generality of metacognition under the unskilled and unaware view. Interestingly, the most exaggerated over-confidence was observed for math, which supports the view that math is unique and that metacognition is domain specific. What is novel about these results is

that students appear to be math anxious yet also overconfident in math. In addition, relative calibration of metacognition for math was generally better than other academic subjects. This overconfidence and greater metacognitive calibration replicated in both studies in this paper, and has also been consistent across our earlier studies (Erickson & Heit, 2013). We have reliably documented over past studies that both high school and college students have over-predicted their scores on math tests, along with biology and literature tests.

Math is indeed unique in some respects and this manifests in an interesting way. Rather than showing lower confidence in math compared to other subjects and poorer metacognition, as would be expected based on the original view that math is unique, students generally showed the highest level of overconfidence in math compared other academic subjects. Students, rather than displaying differing metacognition in the form of lower confidence stemming from math anxiety, instead showed differing metacognition in the form of even more exaggerated overconfidence despite the presence of math anxiety. In Study 1, students were retaking only their Algebra 1 class, not a literature or a biology class. Whereas we did not directly measure math anxiety in this study, it is plausuble that these students were math anxious overall.

Another relatively novel feature of these two studies is the consideration of both predictions and postdictions of performance, rather than just postdictions or individual test item evaluations. Math was unique in that students displayed higher over-confidence when compared to other subjects, and this persisted in both predictions and postdictions. Students also displayed elements of being unskilled and unaware both before and after the test, though their calibration did improve after taking a test. It seems an obvious result that metacognition would improve after students take a test, and it is tempting to treat postdictions as the more relevant measure to be considered when evaluating actual student metacognition. After all, doing so would help in comparing metacognition across students once they have equal footing in knowing exactly what is on the test. However, we would argue that predictions, not postdictions, provide a more realistic and practical measure of student metacognition. Students typically cannot view actual questions before they take a test in an academic setting. Rather, they must use predictions to guide their self-regulation activities, including studying as well as choices such as how to take notes in class (or even whether to attend class). Students' metacognitive skill in postdicting performance after a test might be more accurate, but this cannot help them to improve test performance and academic success.

As noted earlier, our findings are focused on the math-anxious student groups studied here, comparing academic subjects. We would see a comparison to non-math-anxious students as a fascinating but challenging potential topic for future research. Just assembling two groups of students who differ in terms of math anxiety but are equal on other variables would present considerable difficulty. Students who differ in math anxiety likely have other attributes (e.g., demographics, math ability) that also differ. These variables would have to be carefully teased apart in order to make any comparisons of math-anxious and non-math-anxious students.

Possible mechanisms

It was not the purpose of these studies to find the exact mechanisms that underlie metacognitive function for math. In general, finding a similar pattern of results for two different tasks does not necessarily indicate that they are the same mechanistically. However, theories in the math anxiety literature help explain why metacognition for math is unique when compared to other academic domains. Notably, Beilock (2008) along with others (Ramirez, Gunderson, Levine, and Beilock, 2013) has shown that working memory is compromised by math anxiety and also by stressful situations, providing a possible explanation for the reduced metacognitive ability we observed for math. Ashcraft and Krause (2007) further showed that math anxiety and peoples' preoccupation with this fear function as secondary mental tasks that draw on working memory resources necessary for problem solving. Any math problem solving that requires more than simple retrieval of information depends on working memory, so a reduction in working memory capacity can lead to a reduction in math performance. This lower performance then results in a disconnect between typical academic performance and math performance and a corresponding miscalibration in metacognition. People's metacognitive evaluation of math ability may ordinarily be more accurate when they are engaging in tasks in a less stressful environment but becomes less accurate when put under pressure.

So far, we have demonstrated overconfidence as a flaw in self-monitoring about math and other domains. Self-monitoring strategies such as self-testing give learners specific and measureable feedback about how much they know. Without accurate feedback, learners are unable to select appropriate self-regulation strategies to further their learning process (Dunlosky, Hertzog, Kennedy, & Thiede, 2005). Rather, they choose study strategies based on feelings of knowing and judgments of learning, both of which have been shown to be inaccurate self-measures of knowledge (Metcalfe & Finn, 2008). These inaccurate measures lead to student flaws in identifying problem areas of knowledge. Without knowing what they do not understand, learners are unable to make plans to fill in gaps in their knowledge. Students can benefit from using effortful self-monitoring by enacting practice tests at home, where they might not suffer as much detraction from working memory stores. With less stress from a testing environment, they will be able to perform closer to their actual knowledge level. This enables learners to highlight gaps in their knowledge and use this to organize their study time more effectively.

Possible limitations

We do not know to what extent these findings generalize to other populations. Both UC Merced and the local high school are both in the same rural area of California and are not necessarily representative of all learners. Ideally, additional studies will be conducted in other learning settings and with more widely differing populations.

Students in Study 2 were math anxious, and although we did not directly measure math anxiety in Study 1, a later study identified another sample of the same high school population as math anxious using the same MARs measure. From this, we might assume that the students in Study 1 had a similar level of math anxiety. Although we did not include a measure of general anxiety, it is also possible that the populations here had

general anxiety and not just math anxiety. Presence of generally high anxiety would further compromise working memory and, consequently, metacognitive ability.

Although performance was not exactly the same across academic subjects, it does not appear that the exaggerated overconfidence in math was simply result of choosing intrinsically more difficult test items compared to biology and literature. Questions in both studies were normed for difficulty, and the math test in Study 2 was in fact less difficult than either biology or literature, based on test item difficulty ratings provided by the College Board, creators of the standardized test. Furthermore, analyses that included only those questions that participants performed best on still displayed this finding of exaggerated overconfidence. In an effort to improve the realism of this study, a later version of this study was performed in math classes at UC Merced. Students provided predictions and postdictions for class midterms, thus utilizing a much more realistic assessment than an SAT II test taken in a lab setting. Findings generally replicated those from the studies in this paper, so results in this paper were not an artifact of lab setting or the particular assessments used.

Final remarks

Our results are relevant for applications in cognitive science, particularly for studying and improving education. We have seen that students are overconfident in math, yet there is evidence that these same students are math phobic. These views pose two strong deterrents for students to seek practice and improvement in math. If students are overconfident in their mathematical abilities and have anxiety about mathematical tasks, they have little incentive to study the subject. This reluctance likely carries over to other science, technology, and engineering subjects that require a significant amount of math background.

We also know that our use of metacognition does not always lead to calibrated self-views of ability. There are optimal models for allocation of study time, but student behaviors do not conform to these (Son & Sethi, 2006; 2009). Judgments of improvement and learning rate that students use to make time allocation decision are often inaccurate as well (Townsend & Heit, 2010; 2011). Math phobia has come to be so expected that it has started to influence curriculum design. Already changes have been made in computer science programs to deemphasize math (Tucker, Kelemen, & Bruce, 2001) even though math content is fundamental to this area. The spread of this trend to other science, technology and engineering programs would seriously undermine students' foundational math knowledge. Therefore it will be important to develop techniques that improve students' metacognitive calibration for mathematics and other subjects.

We do not doubt that math anxiety exists. However, it is important to differentiate metacognitive judgments of performance from feelings of anxiety, which may have a more emotional or physical, rather than cognitive, basis. That students can be anxious yet overconfident has pernicious implications for struggling math learners. Overconfidence and anxiety provide students with two reasons to avoid studying math or attending math classes. Furthermore, extensive evidence shows that anxiety leads to avoidance (Ashcraft, 2002), implying that anxious students would avoid attending math classes. Other examples of math avoidant behaviors include avoiding lectures, avoiding homework, avoiding study time, or avoiding test preparation. Globally, the 2012

Programme for International Student Assessment (PISA) study, which assessed 15-16 year olds in 65 countries, found that students with higher math anxiety were more likely to have lower math self-concept. Trends in this comprehensive study point toward increases in math anxiety over the past decade.

Metacognition can also further impact other abilities such as attention, memory, perception, comprehension, reasoning, and problem solving (Kitchener, 1983; Metcalfe & Shimamura, 1994) and also affect social behavior (Jaccard, Dodge, & Guilamo-Ramos, 2005) and decision making (Cohen, Freeman, & Thompson, 1998). We have not yet examined the effects of math confidence on math anxiety on self-regulation. Which is a better predictor of study behavior, math confidence or math anxiety? Would pointing out the contradiction between being overconfident about math and being anxious about it have beneficial consequences for struggling students? We see these questions as important for future research.

We finish with a cautionary note. Some educational interventions aim to boost students' math self-confidence, because math self-concept is strongly related to math grades (Marsh, Trautwein, Ludtke, Koller, & Baumert, 2006). It is important to keep in mind that students' confidence in their math performance is probably already high, likely contributing to math avoidance. Aiming to further increase self-confidence in math may have unintended consequences for learning.

Chapter 5: Metacognition, Math Anxiety, and Perceived Rank (Studies 3 & 4)

(Note: Work presented here originally appeared in the Proceedings of the 38th Conference of the International Group of the Psychology of Mathematics Education, 2014)

Studies 1 and 2 showed that metacognitive calibration in math is unaffected by ability or math anxiety. So what is a better predictor? Both ability and math anxiety are self-attributes. However, students do not make self-judgments based solely on internal information. For example, judgments of self-ability are affected by the people with whom we compare ourselves (e.g. Bandura, 1993). Social pressures and expectations have similarly been shown to be a factor in self-perception of ability (e.g. Wood & Bandura, 1989) and as a source of mental distress (Wood, Boyce, Moore, and Brown, 2012). In comparing ourselves to others, we derive comparisons within a relevant group. For example, someone might consider themselves better or worse at certain tasks than the average person. This is simple example of creating a perceived relative comparison ("better" or "worse") within a group of peers. This can also become much more specific. For example, someone might feel they are in the highest or lowest 25% for a certain skill or level of knowledge. Judgments such as this lead to developing a perceived rank among peers.

We now consider student perception of ability rank as a possible influence on metacognitive calibration in math. Individuals' beliefs about their relative ability (rank) within a group of peers can affect decision making, self-judgments, and mental distress. Brown, Gardner, Oswald, and Qian (2008) found that workers' wage rank within a company contributes to pay satisfaction. They argue that hierarchy has a strong influence on people and that it is necessary to consider wage rank within a workplace in addition to simple absolute wage or relative wage when determining worker satisfaction. Wood, Boyce, Moore, and Brown (2012) further present a psychosocial explanation for how income rank predicts mental distress. In general, people with low income are at an increased risk for mental distress. In their 17 year longitudinal study of 30,000 participants, people's income rank served as a type of social rank that strongly influenced their current and future stress. A lower income rank served as a lower perceived social rank, and this social comparison resulted in higher stress. People can also make rank-based decisions for social norms. Wood, Brown, and Maltby (2012) showed that judgments of other people's alcohol consumption influenced drinkers' perception of their own consumption. In their study, people used relative rank rather than absolute measures in making judgments about their own drinking habits. Those who thought of themselves as average or below average drinkers rated themselves at low risk for alcohol-related disorders, regardless of absolute level of alcohol consumption. People further used their perceived rank to make future decisions about their drinking habits. Wood, Brown, and Maltby (2011) also showed that people's gratitude follows a relative rank model. A person's gratitude for an act of help is largely dependent on how this act compares to help the person is used to receiving. People tend to compare individual acts of help with the spectrum of help they have had in the past rather than how much help is actually received (e.g. number of minutes of help).

In order to make metacognitive judgments and choose strategies, students must first make self-judgments about ability. As rank is influential to decision making and

self-judgments, it follows that metacognition is also influenced. If metacognitive calibration is unaffected by math anxiety, is it instead related to how individuals rank their own ability when comparing themselves to a group of peers? We hypothesize that individuals' perceived rank of ability among peers will be associated with metacognitive judgments. In particular, we predict that those who are more accurate in their perceptions of relative rank of ability will also have more accurate metacognitive calibration for self-ability. For example, those who give themselves a mistakenly high rank among their peers might have a correspondingly inflated and inaccurate view of their abilities while those who consider themselves less able than their peers might underrate their actual ability. Furthermore, social rank can influence mental distress, so perceived rank may also affect levels of math anxiety. Specifically, those who rank themselves lower in ability among peers may suffer higher levels of math anxiety, while those who consider themselves more able than their peers may not experience as much math anxiety.

Results from Studies 3 and 4 are below, followed by a general discussion regarding both studies.

Study 3

Method

Participants. Our sample consisted of 53 undergraduate students (40 female, 13 male). All participants were UC Merced undergraduates (*mean age* = 19.11, *sd* = 1.20) who received extra credit in their introductory psychology or cognitive science classes for their participation.

Materials and Procedure. Participants predicted their performance on a math test consisting of 15 sample SAT II test questions released by the College Board. Test questions were normed to avoid ceiling and floor effects. Questions left unfinished were scored as incorrect. Participants also provided predicted distributions for how they thought other participants would perform on the same test. After the test, participants provided a postdicted score. Participants also reported predicted and postdicted scores for the rest of the participants in the study. They were asked to estimate how many people out of 100 participants would score 0 to 1 question correct, 2 to 3 questions correct, etc. They were not told their own scores or others' scores, and they were not able to access their previous predictions when making postdiction judgments. Participants were told that these were based on SAT II Subject Tests (which most students have taken). Then, participants answered 25 questions about math anxiety from the shortened MARS (Alexander & Martray, 1989), using a rating scale ranging from 1 – "no anxiety" to 5 – "very high anxiety", with possible scores from 23 to 125.

Results

Results generally replicated our past studies. Again we see miscalibration in metacognitive judgments. Participants were generally overconfident in predicting their math performance and their calibration improved for postdictions. The regression line for predicted scores vs actual scores shows overconfidence for those with lower performance, while higher performers are more calibrated in their predictions. Students were math anxious overall, with an average adapted shortened MARS score of 72/125 (*sd* = 20.72).

Estimates of peer performance did not change significantly from predictions to postdictions. Mean predicted score for self-judgments of ability was 8.51 (out of 15) while the corresponding mean postdicted score was 6.06. In contrast, the mean predicted score for perceived ability of peers was 8.81 while the corresponding mean postdicted score was 8.34. For reference, the actual mean score over all participants was 6.59. T-tests further show that differences between predictions and postdictions for self-judgements of ability were statistically significant ($t = 2.84$, $df = 156.2$, $p < 0.01$) while differences between predictions and postdictions made for peers were not ($t = 0.684$, $df = 162.0$, $p = 0.50$).

Individuals' perceived rank of their own ability relative to their peers' was highly correlated with metacognitive judgments of their own performance ($r = 0.60$, $p < 0.001$ for predictions; $r = 0.69$, $p < 0.001$ for postdictions). Those who ranked themselves higher among peers tended to be less accurate (over-confident) in their metacognitive calibration, while those who ranked themselves lower among their peers had more accurate self-estimates of performance. Perceived rank was also moderately correlated with performance ($r = 0.39$, $p < 0.01$ for predictions; $r = 0.51$, $p < 0.001$ for postdictions). Those who ranked themselves higher among their peers did also have higher actual scores.

Math anxiety was unrelated to metacognitive calibration but was moderately correlated with math ability ($r = -0.27$, $p < 0.05$). Those with higher math anxiety tended to have lower math ability. Math anxiety was also moderately correlated with how individuals ranked themselves among peers ($r = -0.39$, $p < 0.01$ for predictions; $r = -0.36$, $p < 0.01$ for postdictions). Those who ranked their self-ability lower within a group of peers also tended to have higher math anxiety.

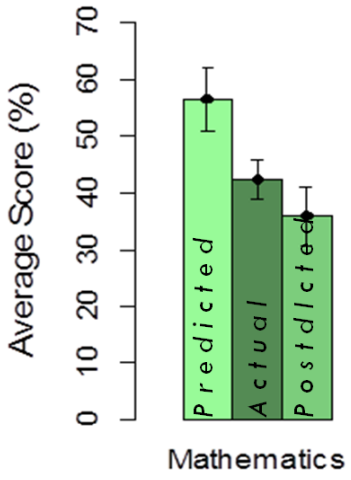

Figure 1: Predictions, actual performance, and postdictions.

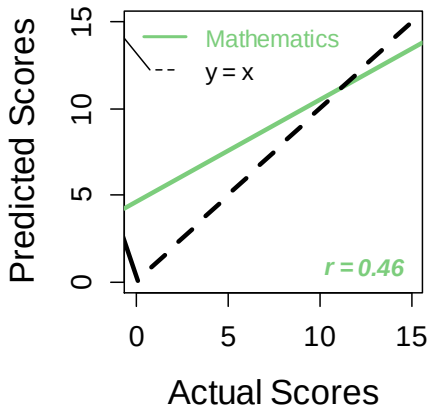

Figure 2: Regression lines for actual versus predicted scores.

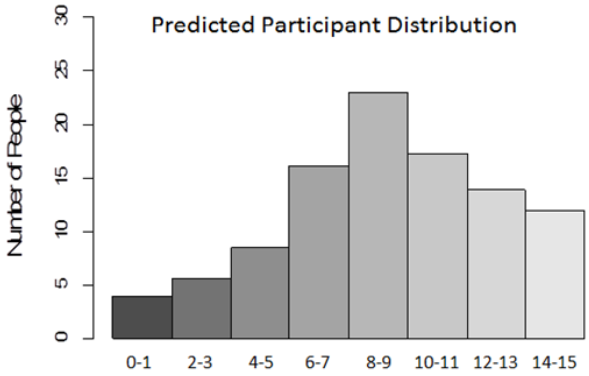

Figure 3: Distribution of Participant-produced Predictions of Peers

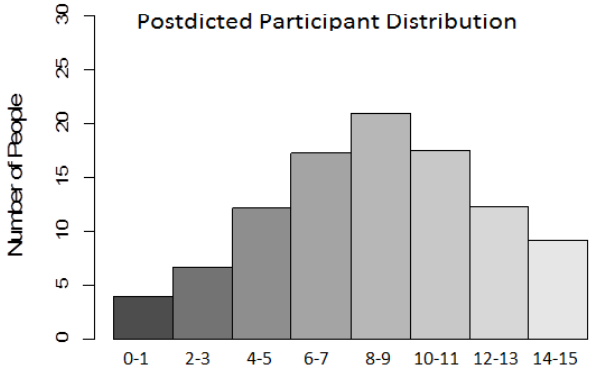

Figure 4: Distribution of Participant-produced Predictions of Peers

Study 4A

Method

Participants. Our sample consisted of 53 undergraduate students (46 female, 37 male). All participants were UC Merced undergraduates (*mean age* = 20.87, *sd* = 1.31) enrolled in Pre-Calculus (Math 5).

Materials and Procedure. Students were informed that their participation was voluntary and would not influence their course grade. Each participant was assigned an anonymous identifier. Students predicted their performance for each of four class midterms during the Spring 2014 semester. As in Study 3, they also provided prediction for how they thought their classmates would perform. After each test, participants provided a postdicted score, as well as postdictions for how they thought their classmates had performed. Actual scores were provided by instructors. Questions required open-ended answers, and number of questions varied for each midterm. Participants also answered 25 questions about math anxiety from the shortened MARS (Alexander & Martray, 1989), using a rating scale ranging from 1 – "no anxiety" to 5 – "very high anxiety", with possible scores from 23 to 125. Students completed the MARS survey online.

Results

Results generally replicated Study 3, despite the change in setting and assessment style. Again we see miscalibration in metacognitive judgments. Participants were generally overconfident in predicting their math performance (mean prediction residual = 8.38%) and their calibration improved for postdictions (mean postdiction residual = 1.25%). The regression line for predicted scores vs actual scores shows overconfidence for those with lower performance, while higher performers are more calibrated in their predictions. Students were math anxious overall, with an average adapted shortened MARS score of 73/125 (*sd* = 31.45).

Estimates of peer performance did not change significantly from predictions to postdictions. Mean predicted score for self-judgments of ability was 75.30% while the corresponding mean postdicted score was 68.18%. In contrast, the mean predicted score for perceived ability of peers was 78.97% while the corresponding mean postdicted score was 77.89%. For reference, the actual mean score over all participants was 66.92%. T-tests further show that differences between predictions and postdictions for self-judgements of ability were statistically significant ($t = 2.84$, $df = 156.2$, $p < 0.01$) while differences between predictions and postdictions made for peers were not ($t = 0.684$, $df = 162.0$, $p = 0.50$).

Individuals' perceived rank of their own ability relative to their peers' was highly correlated with metacognitive judgments of their own performance ($r = 0.79$, $p < 0.001$ for predictions; $r = 0.82$, $p < 0.001$ for postdictions). Those who ranked themselves higher among peers tended to be less accurate (over-confident) in their metacognitive calibration, while those who ranked themselves lower among their peers had more accurate self-estimates of performance. This correlation of metacognitive self-judgments with perceived rank is notably stronger than correlations with either math ability or math anxiety, suggesting that perceived rank is indeed a better predictor of metacognitive ability. Perceived rank was also moderately correlated with performance ($r = 0.48$, $p <$

0.01 for predictions; $r = 0.59$, $p < 0.001$ for postdictions). Those who ranked themselves higher among their peers did also have higher actual scores.

Math anxiety was unrelated to metacognitive calibration but was moderately correlated with math ability ($r = -0.32$, $p < 0.01$). Those with higher math anxiety tended to have lower math ability. Math anxiety was also moderately correlated with how individuals ranked themselves among peers ($r = -0.27$, $p < 0.05$ for predictions; $r = -0.47$, $p < 0.001$ for postdictions). Those who ranked their self-ability lower within a group of peers also tended to have higher math anxiety.

Surprisingly, subsequent midterms did not yield significantly different results from the first midterm. This is perhaps because material for each midterm is considered distinct from other midterms and not as a continuation of the same material. It appears that students reset their expectations for each exam, rather than adjusting (improving) their self-evaluations. We also performed analyses based on gender for both Studies 3 and 4. Differences between these groups proved to be statistically insignificant. Differences we measured include math anxiety, performance, metacognitive self-judgments, peer judgments, and changes from predictions to postdictions.

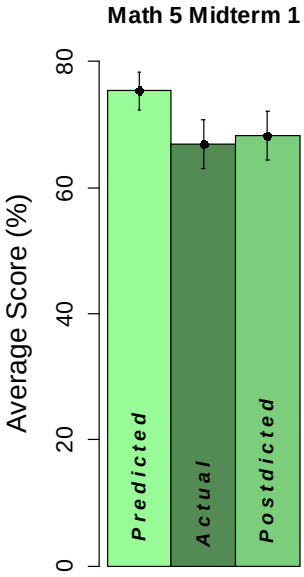

Figure 5: Predicted, actual, and postdicted scores.

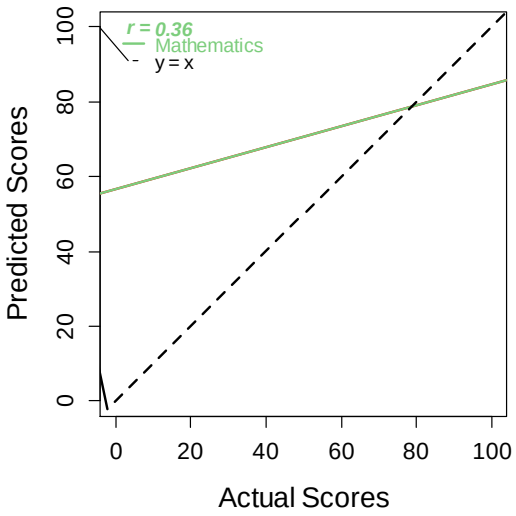

Figure 6: Regression lines for actual versus predicted scores.

Study 4B

Method

Participants. Our sample consisted of 199 undergraduate students (115 female, 84 male). All participants were UC Merced undergraduates (*mean age* = 20.93, *sd* = 1.47) enrolled in Calculus II (Math 12).

Materials and Procedure. Students were informed that their participation was voluntary and would not influence their course grade. Each participant was assigned an anonymous identifier. Students predicted their performance for each of four class midterms during the Spring 2014 semester. They also provided prediction for how they thought their classmates would perform. After each test, participants provided a postdicted score, as well as postdictions for how they thought their classmates had performed. Actual scores were provided by instructors. Questions required open-ended answers, and number of questions varied for each midterm. Participants also answered 25 questions about math anxiety from the shortened MARS (Alexander & Martray, 1989), using a rating scale ranging from 1 – "no anxiety" to 5 – "very high anxiety", with possible scores from 23 to 125. Students completed the MARS survey online.

Results

Results generally replicated Study 3, despite the change in setting and assessment style. Again we see miscalibration in metacognitive judgments. Participants were generally overconfident in predicting their math performance (mean prediction residual = 8.38%) and their calibration improved for postdictions (mean postdiction residual = 1.25%). The regression line for predicted scores vs actual scores shows overconfidence for those with lower performance, while higher performers are more calibrated in their predictions. Students were math anxious overall, with an average adapted shortened MARS score of 87.5 (125 (*sd* = 24.12).

Estimates of peer performance did not change significantly from predictions to postdictions. Mean predicted score for self-judgments of ability was 65.74% while the

corresponding mean postdicted score was 47.43%. In contrast, the mean predicted score for perceived ability of peers was 65.20% while the corresponding mean postdicted score was 59.82%. For reference, the actual mean score over all participants was 48.42%. T-tests further show that differences between predictions and postdictions for both self-judgements of ability and judgements for peers were statistically significant ($t = 8.18$, $df = 396$, $p < 0.0001$; $t = 3.75$, $df = 394.51$, $p < 0.001$).

Individuals' perceived rank of their own ability relative to their peers' was again highly correlated with metacognitive judgments of their own performance ($r = 0.84$, $p < 0.0001$ for predictions; $r = 0.81$, $p < 0.0001$ for postdictions). Those who ranked themselves higher among peers tended to be less accurate (over-confident) in their metacognitive calibration, while those who ranked themselves lower among their peers had more accurate self-estimates of performance. Perceived rank was only weakly correlated with performance for predictions ($r = 0.14$, $p < 0.05$) but was for postdictions ($r = 0.30$, $p < 0.001$ for postdictions). Those who ranked themselves higher among their peers, especially for postdictions, did also have higher actual scores.

Math anxiety was unrelated to metacognitive calibration but was moderately correlated with math ability ($r = -0.27$, $p < 0.0001$). Those with higher math anxiety tended to have lower math ability. Math anxiety was also moderately correlated with how individuals ranked themselves among peers ($r = -0.24$, $p < 0.001$ for predictions; $r = -0.32$, $p < 0.0001$ for postdictions). Those who ranked their self-ability lower within a group of peers also tended to have higher math anxiety. Again subsequent midterms did not yield significantly different results from the first midterm.

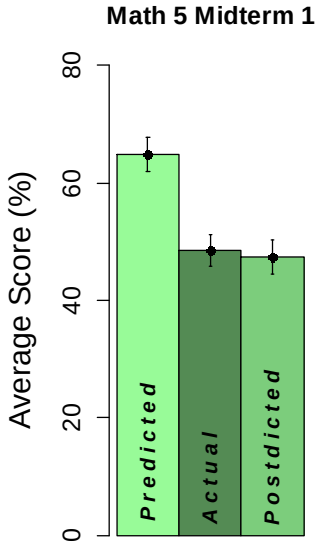

Figure 7: Predicted, actual, and postdicted scores.

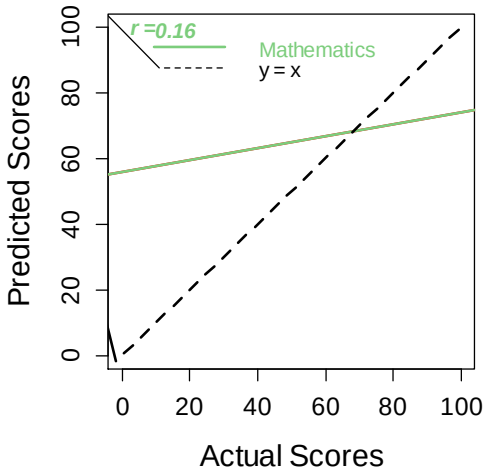

Figure 8: Regression lines for actual versus predicted scores.

General Discussion

Results were very similar for Studies 3 and 4AB, despite fundamental changes in setting and assessment type. Study 3 utilized a multiple choice test given in a lab setting, and participants were given extra credit in their class for their participation. Study 4 was administered in actual math classes using class assessments, and students participated on a voluntary basis. We view this as evidence that supports the general quality of data collected in lab settings, particularly for Studies 1 and 2 in which students had little obvious motivation to perform well on a lab assessment.

In Study 4A, students did not significantly change their estimates of performance for peers from predictions to postdictions, yet students from Study 4B did. This result might be explained by the difference in class level and difficulty. Students in Study 4A were enrolled in a Pre-Calculus course while students in Study 4B were enrolled in a Calculus II course. Interestingly, students in the Calculus II course displayed higher levels of math anxiety than did the Pre-Calculus students. Pre-Calculus students also made more accurate predictions and postdictions of the own performance than did the Calculus II students. As the Calculus II students had higher levels of math anxiety, this might have inhibited their metacognitive ability in making self-judgements of performance in the midst of a stressful math task.

Overall, we conclude that individuals' perceived rank of ability among peers, not actual ability nor math anxiety, could influence metacognitive judgements of math ability, a novel result. So far, perceived rank has shown the strongest link to metacognitive calibration, much more so than either math anxiety or math ability. In our studies, perceived rank, as measured by student-produced distributions, further provides a way to quantify the level of its influence on metacognitive calibration and math anxiety. Participants' dramatic change from predictions to postdictions for their self-judgments of ability was not matched by such a change for their judgments of ability for their peers. In completing the math assessment, participants are recalibrating their perception of their own ability but not the ability of others. Rather, they are holding their initial perceptions of others as given. If students are using rank to inform their metacognitive self-judgments of ability, this can lead to metacognitive miscalibration.

Correlations between perceived rank among peers and metacognitive calibration strongly suggest that these two judgments are influential to each other. However, there is not yet any clear evidence for the direction of this causality. It is possible that perceived rank among peers could be influencing metacognitive judgments of self-ability. For example, a student may over-estimate their skills if they believe themselves to be better able than their classmates. However, metacognitive calibration might instead be influencing perceived rank. This would imply that a student uses self-judgment of ability to rank themselves among peers. As we have seen, supporting work within other domains implies that perceived rank can strongly influence factors such as job satisfaction, stress, and evaluation of one's health habits and risks. For instance, by directly controlling for actual alcohol intake, Wood, Brown, and Maltby (2012) deduced that perceived rank informs alcohol-related risk judgements. This dispelled the possibility that those who rank themselves higher among peers simply also tend to have a higher alcohol intake. From this, it is plausible that the same directionality would hold true within a math task context, in that perceived rank is influencing metacognitive judgments of ability. However, further studies are needed to further define this relationship.

Participants' perceived rank has strong a correlation with metacognitive calibration and a moderate one with math anxiety. We saw from Wood, Boyce, Moore, and Brown (2012) that perception of rank can influence mental distress. Our studies show that student beliefs of their relative rank are associated specifically with math anxiety, another novel result. If students consider themselves to be ranked low within a group of peers, they tend to have higher math anxiety. How students view themselves within the context of a class or a general group of peers has a great impact on metacognitive calibration.

Student perception of rank is also highly relevant to other aspects of education. For example, rank influences to how undergraduates rate their satisfaction. Brown, Wood, Ogden and Maltby (2015) showed that student satisfaction is affected by the context of other institutions. Students will judge satisfaction with their classes, instructors, departments, and institution based on what they believe is true of other universities and how their own institution compares to this. An attribute of an institution (e.g. feedback promptness from instructors) might be rated as poor by a student who believes other institutions are superior in this trait, while another student might rate the same attribute as very high if they believe other universities are inferior in this trait. Such highly varied ratings can be given regardless of absolute measures (e.g. actual feedback times). Student beliefs about other institutions are often incorrect, yet these beliefs are guiding their evaluations of their own institution. Educational institutions in turn can rely on student evaluations in determining policies meant to boost student satisfaction, which may in fact remain constant if evaluations are determined by judgments about other universities.

Further investigations are required for the relationships between perceived rank, metacognition, and academic success. As perceived rank affects metacognition so strongly, students need to be made aware of misconceptions of their rank within a group in order to avoid metacognitive miscalibration as it can be detrimental to student choices in enacting effective study strategies.

Chapter 6: Effects of Psychological Stress on Metacognition (Study 5)

From Studies 1-4, we found perceived rank to be the best predictor so far of metacognitive self-judgments of ability. Those who ranked themselves higher among their peers also gave themselves higher estimates of math ability. Math anxiety, however, was not indicative of metacognitive judgments. It is surprising that such high levels of math anxiety can be present with such overly inflated views of self-ability. As discussed in Chapter 3, self-report measures of stress are often unreliable, as they measure perceived levels of stress. In addition, the survey itself might be suggestive in seeking answers that imply the presence of math anxiety, leading to over-inflated reports of math anxiety levels. Stereotype threat may also be similarly influencing student responses if students believe they are expected to have math anxiety. If math anxiety is considered a form of stress, one could argue that self-report measures of math anxiety, such as MARS, are equally unreliable as general measures of stress.

For the following study, we introduce physical measures stress (blood pressure, heart rate, and salivary cortisol), as well as a self-reported measure of general stress. Psychological stress can interfere with working memory (Duncko, Johnson, Merikangas & Grillon, 2009; Elzinga & Roelofs, 2005; Lupien, Gillin & Hauger, 1999). In particular, it can inhibit spatial working memory (Diamond, Fleshner, Ingersoll, & Rose, 1996; Mizoguchi, Yuzurihara, Ishige, Sasaki, Chui & Tabira, 2000), memory retrieval (Oei, Everaerd, Elsinga, Van Well & Bermond, 2006) as well as working memory-related activity in the dorsolateral prefrontal cortex (Qin, Hermans, van Marle, Luo, & Fernandez, 2009). As metacognitive function depends on working memory, we predict that higher concentrations of cortisol will impair metacognitive judgments of ability. Perceived stress and anxiety do not typically affect salivary cortisol (Eck, Berkhof, Nicolson, & Sulon, 1996), so we also expect MARS scores to differ from cortisol levels as well as other physical measures of stress. Specifically, we predict physical measures of stress to be a better indicator of metacognitive calibration than MARS scores.

Study 5

Method

Participants. Our sample consisted of 29 undergraduate students (20 female, 9 male). All participants were UC Merced undergraduates (*mean age* = 20.45, *sd* = 2.75) who received extra credit in their introductory psychology or cognitive science classes for their participation.

Materials and Procedure. Participants predicted their performance on a math test consisting of 15 sample SAT II test questions released by the College Board. Test questions were normed to avoid ceiling and floor effects. Questions left unfinished were scored as incorrect. Participants also provided predicted distributions for how they thought other participants would perform on the same test. After the test, participants provided a postdicted score. Participants also reported predicted and postdicted scores for the rest of the participants in the study. They were asked to estimate how many people out of 100 participants would score 0 to 1 question correct, 2 to 3 questions correct, etc. They were not told their own scores or others' scores, and they were not able to access their previous predictions when making postdiction judgments. Participants were told that these were based on SAT II Subject Tests (which most students have

taken). Participants also answered 25 questions about math anxiety from the shortened MARS (Alexander & Martray, 1989), using a rating scale ranging from 1 – "no anxiety" to 5 – "very high anxiety", with possible scores from 23 to 125.

Cortisol concentrations are highest in the morning and decrease throughout the day, reaching almost zero at night. Participant sessions were scheduled between 1pm and 5pm, a span of time that minimizes circadian fluctuation of cortisol concentrations during the study and minimizes variation in cortisol concentrations across participants. Participants were instructed not to eat or drink anything except water in the 30 minutes prior to the study. They were also instructed not to engage in rigorous exercise in the 12 hours prior to the study. After granting informed consent, participants were screened for use of medications that contain corticosteroids (e.g. inhaler, asthma medication, hydrocortisone) and medications for treating depression. These participants and participants who smoked were removed from the study. Measurements for height, weight, body fat percentage, water percentage, bone percentage, and muscle percentage were taken for the remaining participants. Participants were then seated at a computer to fill out demographic information (age, gender, major). Baseline measurements (time point 1) were collected for heartrate, blood pressure, and cortisol concentration. Participants also filled out a baseline subjective stress test survey. After this, participants completed the shortened MARS math anxiety ratings scale survey. The stressor item (math test) followed this. Another set of measurements for heartrate, blood pressure, salivary cortisol, and subjective stress was collected 10 minutes after the start of the stressor (time point 2). Participants were allowed 18 minutes to finish the 15-question math test. After the math test, participants were seated in another room and were told to relax for the rest of the study. They were also provided with optional reading material (e.g. magazines). Additional measures for heartrate, blood pressure, salivary cortisol, and subjective stress were then taken at 20 minutes and 40 minutes (time points 3 and 4) past the onset of the stressor.

Cortisol concentrations were derived from saliva samples. For each of the four time point measurements, participants were instructed to place a cotton swab under their tongue for two minutes. These cotton swabs were then placed in salivette tubes. Samples were frozen and stored within one hour of collection until analyses for cortisol measurement were conducted at a later time.

Cortisol present in saliva samples was measured using enzyme-linked immunosorbent assay (ELISA) kits that employ a horseradish perozidase labelled cortisol tracer. Before measurement analyses, samples were thawed and centrifuged. In a salivette, cotton swabs are contained in an inner basket tube, and this basket is contained in an outer polypropylene tube. Centrifuging removes saliva from the cotton swabs into the outer polypropylene tube. Cotton swabs and baskets are then discarded, and any saliva remaining after analyses are refrozen in the outer polypropylene tubes. For each ELISA kit, cortisol standards (3.0, 1.0, 0.333, 0.111, 0.037, 0.012 ug/dL), controls, and saliva samples were pipetted into a 96-well microtitre plate, and these compete with cortisol conjugated to horseradish peroxidase for antibody binding sites on the plate. Samples were incubated, then unbound components were removed using a plate washer. Bound cortisol enzymes conjugate was measured by the reaction of the horseradish peroxidase enzyme to a substrate (tetramethylbenzidine, TMB). The reaction was

stopped using an acidic solution (sulfuric acid), and optical density was read on a plate reader at 450nm to measure cortisol enzyme conjugate. The cortisol present was calculated as an inverse proportion of the cortisol enzyme conjugate measured.

Results and Discussion

Results mostly replicate Studies 1-4 but with some departures. We see recurring miscalibration in metacognitive judgments. Participants were generally overconfident in predicting their math performance (mean prediction residual = 22.07%) and their calibration improved for postdictions (mean postdiction residual = –1.15%). The regression line for predicted scores vs actual scores shows the most exaggerated overconfidence we have yet reported for those with lower performance, while higher performers become underconfident in their predictions. This can also been seen in the strong correlations between ability and calibration of predicted and postdicted performance. Test scores were negatively correlated with accuracy of score estimation ($r = -0.59$, $p < 0.001$ for predictions; $r = -0.44$, $p < 0.05$ for postdictions). Those with higher scores were less confident in their ability. This is surprising given that the overall performance on the math test was only 27.82%, which is arguably a snapshot of only low performers if performance can be compared to previous studies using the same assessment piece.

Estimates of peer performance did follow similar patterns to previous studies and did not change significantly from predictions to postdictions. Mean predicted score for self-judgments of ability was 49.89% while the corresponding mean postdicted score was 26.67%. In contrast, the mean predicted score for perceived ability of peers was 56.56% while the corresponding mean postdicted score was 48.98%. For reference, the actual mean score over all participants was 27.82%. T-tests do show that differences between predictions and postdictions for both self-judgements of ability and judgements for peers were statistically significant ($t = 5.19$, $df = 50.12$, $p < 0.0001$; $t = 2.55$, $df = 52.62$, $p < 0.05$).

Individuals' perceived rank of their own ability relative to their peers' continued to show a stable high correlation with metacognitive judgments of their own performance ($r = 0.81$, $p < 0.0001$ for predictions; $r = 0.56$, $p < 0.01$ for postdictions). Those who ranked themselves higher among peers tended to be less accurate (over-confident) in their metacognitive calibration, while those who ranked themselves lower among their peers had more accurate self-estimates of performance. However, unlike in previous studies, perceived rank was not correlated with performance either for predictions or postdictions. This indicates that students who rank themselves higher among peers are not doing so simply because they are also high performers.

Students were math anxious overall, with an average adapted shortened MARS score of 71.97/125 ($sd = 18.60$). This level of math anxiety is comparable to that of participants in the previous studies. In contrast with previous studies, math anxiety was related to metacognitive calibration for predictions ($r = – 0.44$, $p < 0.05$), though not postdictions. In addition, math anxiety was not correlated with math ability. Math anxiety was again correlated with how individuals ranked themselves among peers ($r = – 0.48$, $p < 0.01$) for predictions but not for postdictions. Those who ranked their self-ability lower within a group of peers also tended to have higher math anxiety. Math

anxiety was also correlated with general confidence when making predictions of performance ($r = -0.53$, $p < 0.01$). Higher math anxiety tended to be linked with lower predictions of performance.

Students showed varied response to that math test as a stressor item (see Figures 3 and 4). Twelve participants showed an increased in cortisol by the end of the stressor, and the 17 other participants did not show any increase in cortisol by the end of the stressor. To determine this, changes in cortisol were measured from time point 1 (baseline) to time point 3 (2 minutes after the end of the stressor and 20 minutes after the onset of the stressor). Cortisol levels were expected to taper after the end of the stressor, as is typical of stress responses, and we did observe this pattern. Non-responders showed no apparent HPAA activation in response to the stressor. Their cortisol levels generally decreased over the course of the experiment, which reflects a pattern typically seen on days without any marked stressors or related HPAA activation. Responders showed increases in cortisol after the onset of the stressor that last through the end of the stressor. Measures of heart rate and blood pressure reveal results similar to that of cortisol through the end of the stressor duration.

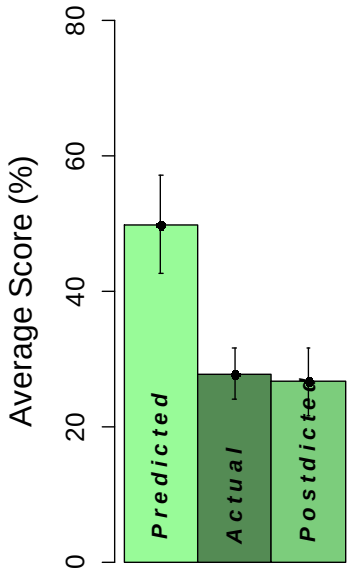

Figure 1: Predicted, actual, and postdicted scores.

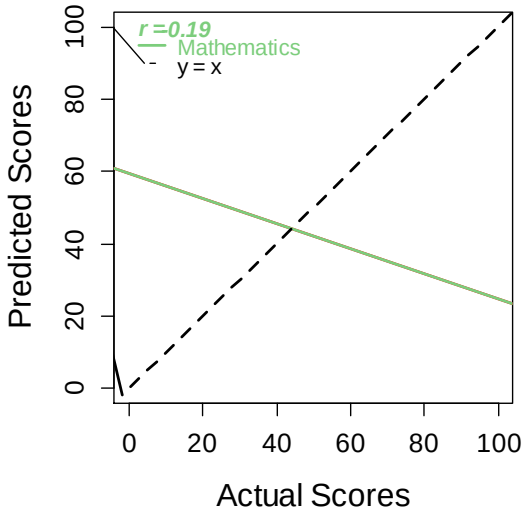

Figure 2: Regression lines for actual versus predicted scores.

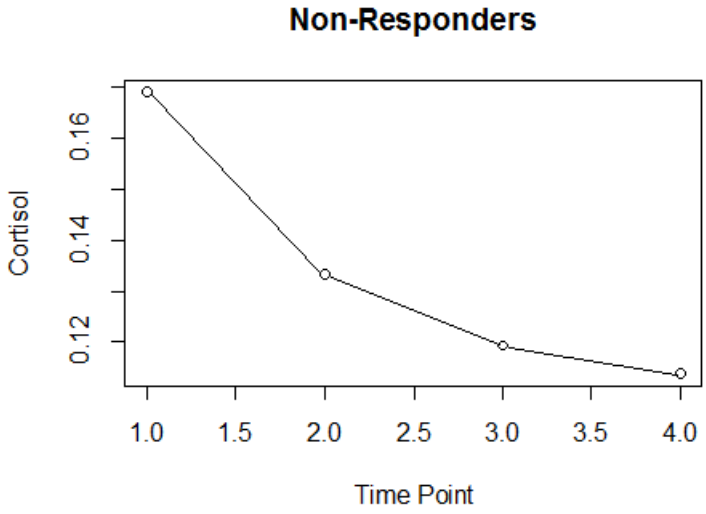

Figure 3: Timeplot of salivary cortisol over four time points for non-responders.

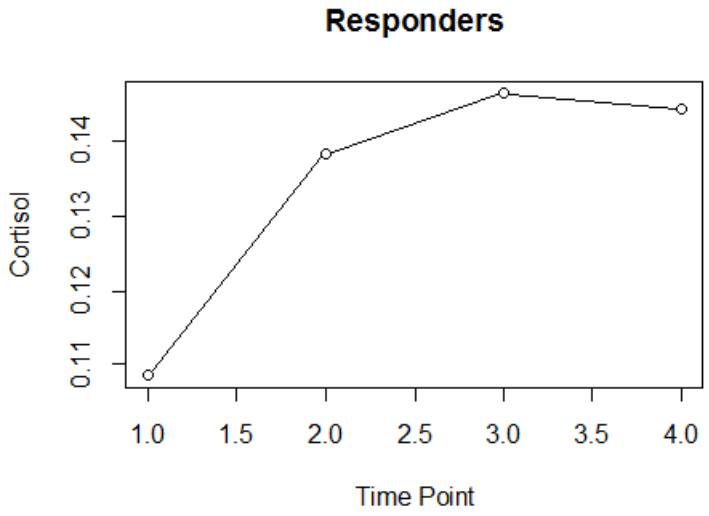

Figure 4: Timeplot of salivary cortisol over four time points for responders.

The average MARS score for responders and non-responders were 62.78 and 76.10, respectively, while cortisol concentrations at time point 3 were 0.146ug/dL and 0.119ug/dL. This shows that participants who experienced HPAA triggers did not necessarily have correspondingly high math anxiety. Similarly, those with high reported math anxiety did not necessarily experience HPAA activation. Overall, MARS scores and cortisol concentrations were unrelated. MARS scores and cortisol concentrations were correlated specifically for non-responders ($r = -0.53$, $p < 0.05$), while not for responders. For responders, cortisol concentrations were highly indicative of calibration for metacognitive judgements of ability ($r = -0.75$, $p < 0.05$). Those with higher cortisol levels were more likely to make more accurate estimates of performance. For non-responders, cortisol was not related to metacognitive calibration. Cortisol was similarly related to judgments of rank among peers ($r = -0.84$, $p < 0.01$). Those with higher cortisol levels also tended to rank themselves lower among their peers for math performance. This is not surprising, as we have previously found perceived rank to be linked with metacognitive calibration. Cortisol levels were unrelated to actual performance for both groups. Similar to MARS scores, Subjective Stress Test scores were lower for responders (44.13 out of 144) and higher (46.91) for non-responders. However, MARS scores were unrelated to Subjective Stress scores, providing evidence that MARS does measure something other than general stress. Subjective Stress Test scores were also unrelated to cortisol levels and metacognitive judgments of ability.

General Discussion

Results were largely similar to studies Studies 3 and 4 but with a few departures. While there was overconfidence in the lower half of performers, the upper half showed significant overconfidence that did not appear in previous studies. Math ability was unrelated to math anxiety. Math test scores were particularly low, yet math anxiety scores were not. These two measurements have already shown an inconsistent

relationship over past studies. However, math anxiety was correlated with general confidence in making predictions of performance. In addition, math anxiety was related to metacognitive calibration, unlike findings from our previous studies. The relationship between rank and metacognitive calibration does remain, indicating a more stable connection than metacognitive calibration with other variables such as math anxiety. Cortisol proved to be another strong indicator for metacognitive calibration, a new result within the intersection of stress and metacognition research. We saw the math test did not cause HPAA activation and resulting physical evidence of stress in all students, which is too be expected – not everyone fears math. However, it is puzzling that students experiencing high levels of stress when exposed to a math test stressor could report lower MARS scores than those students who did not have an HPAA stress reaction to the same math test. Similarly, those with a low HPAA stress reaction did not report correspondingly low MARS scores. From this we deduce that MARS scores measure something other than psychological stress caused by HPAA activation. The Subjective Stress Test was similarly unrelated to HPAA activation, and this is consistent with previous findings for the lack of connection between subjective stress tests and cortisol levels (e.g. Hellhammer, Wust, & Kudielka, 2009; Van Eck, Berkhof, Nicolson, & Sulon, 1996). This is not surprising, as these tests can be based on varied theoretical constructs. Low associations of MARS and Subjective Stress tests may also result from neuroendocrine factors. While HPAA activation does lead to increased cortisol production, other influences such as the hippocampus, hypothalamus, pituitary, and adrenals may also be affecting cortisol measures.

Our findings weaken the potential usefulness of these self-report measures. If math anxiety is thought to induce fear and stress, then perhaps MARS is not measuring math anxiety at all. Rather, it might be measuring a perceived construct of what students believe their math anxiety should be rather than the math anxiety they actual experience. This would help to explain the inconsistencies for the relationship between math anxiety and ability and between math anxiety and metacognitive calibration. As some people respond to math as a visceral threat or even with actual experience of pain, there is debate whether math anxiety is only psychological (Beilock, 2012). This pain network activation was only present in anticipation of math, however, and not the actual math task itself. It is possible that non-responders classified in Study 5 reacted to their anticipation of the stressor, rather than the stressor itself, followed by a decrease in HPAA activation from the start of the stressor.

Our overall conclusion remains that individuals' perceived rank of ability among peers is more reliably influential to metacognitive judgments of ability than are math ability or math anxiety. In addition, cortisol and other physical measures of stress are promising new variables that may aid in determining metacognitive ability for math tasks. These variables have not been used in and previous studies of metacognition.

Chapter 7: Summary, Overall Conclusions, and Future Directions

We have seen that metacognition plays a role in cognitive processing that is particularly important for educational pursuits. It impacts several areas within cognitive science, and aids complex thought processes necessary for intellectual endeavors, and plays fundamental role in forming self-judgments of ability and knowledge which are of particular importance in educational contexts.

Several components of cognition affect the learning of mathematical concepts and procedures, and mathematical cognition can be broken into smaller components of varying complexity. The multidisciplinary nature of mathematical cognition emphasizes the importance of understanding mathematical cognition within many domains, from industry to education, and highlights new issues surrounding mathematical skills with implications across research and education. Understanding math anxiety is of particular importance. Math anxiety can negatively impact cognition in a variety of ways. These can be seen through the impact on working memory, and the resulting impact on performance. Negative effects of math anxiety, as a form of psychological stress, can also be seen through the measurement of cortisol. Cortisol arguably provides a more accurate measure of psychological stress induced by math tasks than do self-reported perceived stress or anxiety tests. High levels of cortisol are associated with decreased cognitive function, particularly working memory, and this might indicate that cortisol could serve as a measure of metacognitive function as well.

In Studies 1 and 2, we measured metacognitive judgments and performance across biology, literature, and math content. Undergraduates took three shortened SAT II Subject Tests, and provided estimates of their performance both before and after taking each test. We found that judgments differed across both domain and gender. Overconfidence was evident in all three domains, although estimates of ability were more accurate after taking a test rather than before. Improvement in calibration and gender differences in calibration were most noticeable in math. Males tended to be more overconfident, while females were less confident yet more accurately calibrated when estimating ability. Students were generally over-confident in math, and this brought into question the existence of math phobia.

From this, our basic conclusion was that students aligned in some ways to both the unskilled and unaware phenomenon and the idea that math is unique in some way. In these first two studies, students over-estimated their performance in their predicted scores for all domains, though their calibration did improve for postdictions. This provided support for the domain generality of metacognition under the unskilled and unaware view. The most exaggerated over-confidence was observed for math, surprisingly, and this supports the view that math is unique and that metacognition is domain specific. This led to the novel and unexpected finding that students studied are math anxious yet also overconfident in math. However, relative calibration of metacognition for math was generally better than other academic subjects.

We concluded that math is indeed unique in some respects. Rather than showing lower confidence in math compared to other subjects and poorer metacognition, as would be expected based on the original view that math is unique, students generally showed the highest level of overconfidence in math compared other academic subjects. Students did not exhibit differing metacognition in the form of lower confidence stemming from math

anxiety. Instead, they showed differing metacognition in the form of even more exaggerated overconfidence despite the presence of math anxiety.

Studies 1 and 2 consider both predictions and postdictions of student performance. While students showed a general ability to improve their calibration for postdictions, it is more realistic to consider prediction calibration as something that drives student study behavior. As such, we consider prediction calibration to be the more important of the two measures.

Until this point, we demonstrated overconfidence as a flaw in self-monitoring about math and other domains. Self-monitoring strategies such as self-testing give learners specific and measureable feedback about how much they know. Without accurate feedback, learners are unable to select appropriate self-regulation strategies to further their learning process, and instead they choose study strategies based on feelings of knowing and judgments of learning, both of which can be inaccurate self-measures of knowledge. These inaccurate measures then lead to student flaws in identifying problem areas of knowledge. Without knowing what they do not understand, learners are unable to make plans to fill in gaps in their knowledge. If learners can highlight gaps in their knowledge, they can organize their study time more effectively.

Results from Studies 1 and 2 show that students are overconfident in math, yet there is evidence that these same students are math phobic. These views pose two strong deterrents for students to seek practice and improvement in math. This reluctance likely carries over to other science, technology, and engineering subjects that require a significant amount of math background. Math phobia has also come to be so expected that it has started to influence curriculum design that de-emphasize math concepts in classes and math requirements in general. The spread of this trend to other science, technology and engineering programs would seriously undermine students' foundational math knowledge. Therefore it will be important to develop techniques that improve students' metacognitive calibration for mathematics and other subjects.

Despite some unexpected findings from Studies 1 and 2, this did not lead us to doubt the existence of math anxiety. From this, however, believe it is important to differentiate metacognitive judgments of performance from feelings of anxiety, which may have a more emotional or physical, rather than cognitive, basis.

Studies 3 and 4 introduce perceived rank of self-ability among peers as an additional variable that could contribute to metacognitive calibration for math tasks. Results were very similar for Studies 3 and 4, despite fundamental changes in setting and assessment type. Study 3 utilized a multiple choice test given in a lab setting, and participants were given extra credit in their class for their participation. Study 4 was administered in actual math classes using class assessments, and students participated on a voluntary basis. This provides support for the general quality of data collected in lab settings, particularly for Studies 1 and 2 in which students had little obvious motivation to perform well on a lab assessment.

In Study 4A, students did not significantly change their estimates of performance for peers from predictions to postdictions, yet students from Study 4B did. This result might be explained by the difference in class level and difficulty. Students in Study 4A were enrolled in a Pre-Calculus course while students in Study 4B were enrolled in a Calculus II course. Students in the Calculus II course displayed higher levels of math

anxiety than did the Pre-Calculus students. Pre-Calculus students also made more accurate predictions and postdictions of the own performance than did the Calculus II students. Calculus II students had higher levels of math anxiety, and this might have negatively impacted their metacognitive ability in making self-judgements of performance while engaging in a stressful math task.

Overall, we concluded from Studies 3 and 4 that individuals' perceived rank of ability among peers, not actual ability nor math anxiety, could influence metacognitive judgements of math ability. Perceived rank in these studies showed the strongest link to metacognitive calibration, much more so than either math anxiety or math ability. Perceived rank, as measured by student-produced distributions, provides a way to quantify the level of its influence on metacognitive calibration and math anxiety. Differences in participants' predictions and postdictions for their self-judgments of ability was not matched by such a change for their judgments of ability for their peers. In completing the math assessment, participants recalibrated their perception of their own ability but not the ability of others, and they are holding their initial perceptions of others as given. If students are using rank to inform their metacognitive self-judgments of ability, this can lead to metacognitive miscalibration.

Participants' perceived rank had strong a correlation with metacognitive calibration and a moderate one with math anxiety. We further observed that student beliefs of their relative rank are associated specifically with math anxiety. If students consider themselves to be ranked low within a group of peers, they tend to have higher math anxiety. How students view themselves within the context of a class or a general group of peers had a great impact on metacognitive calibration.

Study 5 includes physical measures of stress that will results from HPAA activation, areas of the brain used for processing fear and stress. Results from Study 5 were largely similar to Studies 3 and 4. However, while there was overconfidence in the lower half of performers, the upper half showed significant overconfidence that did not appear in previous studies. Math ability was unrelated to math anxiety. Math test scores were particularly low, yet math anxiety scores were not. These two measurements have already shown an inconsistent relationship over past studies. Math anxiety was correlated with general confidence in making predictions of performance and was also correlated with metacognitive calibration. The relationship between rank and metacognitive calibration was consistent with Studies 1-4, and this indicated a more stable connection than metacognitive calibration with other variables such as math anxiety. Cortisol was strong indicator for metacognitive calibration, a new result within the intersection of stress and metacognition research. We saw that math does not necessarily cause HPAA activation and resulting physical evidence of stress in all students. This makes sense, as not everyone fears math. We saw that students experiencing high levels of stress when exposed to a math test stressor could report lower MARS and Subjective Stress Test scores than those students who did not have an HPAA stress reaction to the same math test. Also, those with a low HPAA stress reaction did not report correspondingly low MARS and Subjective Stress Test scores. This indicates that MARS scores could be measuring something other than actual stress. These findings weaken the usefulness of these self-report measures. If math anxiety is thought to induce fear and stress, then MARS may not be measuring math anxiety at all. This

helps explain the inconsistencies for the relationship between math anxiety and ability and between math anxiety and metacognitive calibration.

Across Studies 1-5, our overall conclusion is that individuals' perceived rank of ability among peers is more reliably influential to metacognitive judgments of ability than are math ability or math anxiety. In addition, cortisol is a promising new variable that may aid in determining metacognitive ability for math tasks.

Future Directions

Both UC Merced and the local high school are both in the same rural area of California and are not necessarily representative of all learners. Thus, we do not know to what extent our findings generalize to other populations. Ideally, additional studies should be conducted in other learning settings and with more widely differing populations.

We have seen that metacognition can also further impact other abilities such as attention, memory, perception, comprehension, reasoning, and problem solving and also affect social behavior and decision making. We highlighted the need to examine the effects of math confidence on math anxiety on self-regulation. It is unclear whether math confidence or math anxiety would be a better predictor of study behavior, and we see these questions as important for future research.

Further investigations are also required for the relationships between perceived rank, metacognition, and academic success. As perceived rank affects metacognition so strongly, students need to be made aware of misconceptions of their rank within a group in order to avoid metacognitive miscalibration as it can be detrimental to student choices in enacting effective study strategies.

Correlations between perceived rank among peers and metacognitive calibration strongly suggest that these two judgments are influential to each other. It is possible that perceived rank among peers could be influencing metacognitive judgments of self-ability, but there is not yet any clear evidence for the direction of this causality. Further studies that control for ability and calibration are needed to further define this relationship.

Broader Implications

These findings hold implications for both students and educators. Student achievement in math ultimately depends on their learning process. If this process is misguided by inaccurate beliefs about ability, students will not reach their full potential in learning math. Furthermore, educators who are not aware of metacognitive calibration in students may inadvertently exacerbate the problem by bolstering the confidence of math anxious students who are likely to already have an inflated view of their math ability. Educators will first need to understand student metacognition and influences on metacognition.

We suggest physical measures of psychological stress as variables that can influence metacognitive calibration. Physical measures of stress, including cortisol, have previously not been used in studying metacognition, though stress is acknowledged to have significant influence on many mental processes related to metacognition. We see this novel approach as a promising intersection of research disciplines. However, while psychological stress response might provide a better predictor for metacognitive ability,

math anxiety remains an important issue in the literature and its presence should not be dismissed. Both students and educators should be aware of the possible influences math anxiety can have on math ability and metacognitive calibration. However, it remains to be seen what math anxiety surveys actually measure in the form of self-report responses. Results from these surveys are not an accurate predictor of human psychological stress response, and its presence does not necessarily affect metacognitive calibration or math ability.

We have seen that students are largely inaccurate in assessing their math ability. This leads to inefficient and ineffectual study choices that will further negatively impact math achievement and career choices. We have not yet discussed how to improve metacognitive miscalibration. This will be an important next step in the literature. Interventions will almost certainly require educators to aid students in learning about their metacognition and what can influence it. One simple way to put students on a better path would be to make them aware of this tendency for overconfidence. This simple act could help them avoid this pitfall in their educational pursuits. Students can also benefit from further guidance in determining gaps in their math knowledge and ability. Such guidance could take the form of suggested self-test strategies. For example, while students might look over a math problem and believe they understand and can replicate the solution process, they should be encouraged to try problems independently to determine their ability. If students are using perceived rank in assessing their own math ability, it would be helpful for teachers to provide overall class performance information to avoid student misconceptions about how their peers are performing.

Miller and Geraci (2011) have shown that metacognition can be improved, but measured improvement was limited to metacognitive monitoring and did not extend to control. In their experiments, incentives such as extra credit as well as detailed feedback on predictions of performance improved students' accuracy in predicted exam performance. However, actual exam scores did not increase, and any improvement in metacognitive monitoring in the form of performance prediction was limited to low-performing students. Higher performing students displayed no difference in calibration or performance. However, there is evidence that metacognitive training can benefit all ability levels when used in combination with cooperative work (Kramarski & Mevarech, 2003).

The simple act of having students predict their score on an exam can benefit metacognitive calibration, and this has other pedagogical value as well. Miller and Geraci note that it teaches students not just to evaluate the scope of a class, exam, or assignment. Rather, this encourages students to consider their preparedness, which possibly increases students' sense of responsibility for their performance in a class. While metacognition is now commonly discussed in educational settings, there is no strong set of literature that yet guides teachers in incorporating metacognition into school curriculum. There already exist metacognitive training methods that have shown improved metacognitive knowledge for students, such as IMPROVE (Mevarech & Kramarski, 1997a), that teach self-directed questioning techniques for comprehension, strategy selection, and making connections within the material. Students who were trained using IMPROVE showed increased metacognitive knowledge as well as improved student performance in math. However, effects of metacognitive training have

not been tracked over time. We do not know if students' improved metacognitive knowledge and performance carry over to future math content and classes.

Furthermore, for these methods to be effective at all, teachers much themselves be aware of metacognitive improvement strategies and know how to implement them. This is something that is not systematically included in teacher training curriculums. Teachers must provide strategic instruction and have the ability to be flexible in their feedback to students. This demanding process requires teachers to be aware of the strategies they are trying to convey and also instructional strategies they themselves can use, and current teacher training programs do not prepare teachers for this type of teaching.

Metacognition remains one of the most promising areas of research within an educational context. Our results are relevant for applications in cognitive science, particularly for studying and improving mathematics education. A metacognitive approach to mathematics education fits well with the number of fundamental components of metacognition that are readily applicable to student tasks (e.g. monitoring and regulating cognition). Through further study we can hope to better understand how we conceptualize metacognition within mathematics education, how we can more accurately assess and predict metacognition, and perhaps eventually we can find ways to develop metacognition.

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive science*, *26*(2), 147-179.

Alexander, L., & Martray, C. R. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in counseling and development*.

Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: Directions for future research. *Developmental review*, *15*(1), 1-37.

Anderson, J. R., Lebiere, C., Lovett, M., & Reder, L. (1998). ACT-R: A higher-level account of processing capacity. *Behavioral and Brain Sciences*,*21*(06), 831-832.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Anderson, J. R. (1990). *Cognitive psychology and its implications*. WH Freeman/Times Books/Henry Holt & Co.

Anderson, J. A. (1993). *Neurocomputing* (Vol. 2). MIT press.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The journal of the learning sciences*, *4*(2), 167-207.

Anderson, J. R., Qin, Y., Sohn, M. H., Stenger, V. A., & Carter, C. S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review*, *10*(2), 241-261.

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child development*, 695-701.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of experimental social psychology*, *35*(1), 29-46.

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition & Emotion*,*8*(2), 97-125.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic bulletin & review*, *14*(2), 243-248.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, *27*(3), 197-205.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of experimental psychology: General*, *130*(2), 224.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, *2*(3), 213-236.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1), 75-106.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current directions in psychological science*, *11*(5), 181-185.

Ashcraft, M. H., Kirk, E. P., & Hopko, D. (1998). On the cognitive consequences of mathematics anxiety.

Baillargeon, R. (1995). A model of physical reasoning in infancy. In *In C. Rovee-Collier, & LP Lipsitt (Eds.), Advances in infancy research*.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist*, *28*(2), 117-148.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive psychology*, *60*(1), 40-62.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education*, 141-157.

Barrouillet, P., Fayol, M., & Lathulière, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning difficulties. *International Journal of Behavioral Development*, *21*(2), 253-275.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201-221.

Bassett, J. R., Marshall, P. M., & Spillane, R. (1987). The physiological measurement of acute stress (public speaking) in bank employees. *International Journal of Psychophysiology*, *5*(4), 265-273.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail working memory and "choking under pressure" in math. *Psychological Science*,*16*(2), 101-105.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*(5), 339-343.

Beilock, S. (2010). *Choke: What the secrets of the brain reveal about getting it right when you have to*. Simon and Schuster.

Benjamin, A. S., Bjork, R. A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterization of the utility of subjective fluency. *Acta Psychologica*, *98*(2), 267-290.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*(3), 615-634.

Bidjerano, T. (2005). Gender Differences in Self-Regulated Learning. *Online Submission*.

Bisanz, G. L., Vesonder, G. T., & Voss, J. F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning: A developmental study. *Journal of Experimental Child Psychology*, *25*(1), 116-128.

Boaler, J. (2008). Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach. *British Educational Research Journal*, *34*(2), 167-194.

Borkowski, J., & Muthukrishna, N. (1992). Moving metacognition into the classroom:" Working models" and effective strategy teaching.

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*(3), 223-240.

Brannon, E. M. (2006). The representation of numerical magnitude. *Current opinion in neurobiology*, *16*(2), 222-229.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational researcher*, 14-21.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of personality and social psychology*, *76*(2), 246.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, 65-116.

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American psychologist*, *52*(4), 399.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being? *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355-389.

Brown, G. D., Wood, A. M., Ogden, R. S., & Maltby, J. (2015). Do student evaluations of university reflect inaccurate beliefs or actual experience? A relative rank model. *Journal of behavioral decision making*, *28*(1), 14-26.

Brysbaert, M. (2005). Number recognition in different formats. *Handbook of mathematical cognition*, 23-42.

Burbaud, P., Camus, O., Guehl, D., Bioulac, B., Caillé, J. M., & Allard, M. (1999). A functional magnetic resonance imaging study of mental subtraction in human subjects. *Neuroscience letters*, *273*(3), 195-199.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281.

Butterworth, B. (1999). What counts: How every brain is hardwired for math. *New York: Free Press.*

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

Canfield, R. L., & Smith, E. G. (1996). Number-based expectations and sequential enumeration by 5-month-old infants. *Developmental Psychology*, *32*(2), 269.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia*, *39*(11), 1224-1239.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and onto genesis. *Mind & Language*, *16*(1), 37-55.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438-448.

Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology*, *8*(6), 583-595.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child development*, 11-28.

Cen, H., Koedinger, K., & Junker, B. (2005). Automating cognitive model improvement by A* search and logistic regression. In *Proceedings of AAAI 2005 Educational Data Mining Workshop*.

Chochon, F., Cohen, L., Van De Moortele, P. F., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of cognitive neuroscience*, *11*(6), 617-630.

Cohen, L., & Dehaene, S. (1996). Cerebral networks for number processing: Evidence from a case of posterior callosal lesion. *Neurocase*, *2*(3), 155-174.

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science*, *5*(2), 186-201.

Cohen, M. S., Freeman, J. T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. *Making decisions under stress: Implications for individual and team training*, 155-190.

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2011). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities*, 0022219410393012.

Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*(4), 354.

Cook, N. J., Read, G. F., Walker, R. F., Harris, B., & Riad-Fahmy, D. (1986). Changes in adrenal and testicular activity monitored by salivary sampling in males throughout marathon runs. *European journal of applied physiology and occupational physiology*, *55*(6), 634-638.

Cooper, T. R., Trunkfield, H. R., Zanella, A. J., & Booth, W. D. (1989). An enzyme-linked immunosorbent assay for cortisol in the saliva of man and domestic farm animals. *Journal of Endocrinology*, *123*(2), R13-R16.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. *Handbook of humancomputer interaction*, 849-874.

Corbett, A., McLaughlin, M., Scarpinatto, K. C., & Hadley, W. (2000, January). Analyzing and generating mathematical models: an algebra II cognitive tutor design study. In *Intelligent Tutoring Systems* (pp. 314-323). Springer Berlin Heidelberg.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395-438.

Coutinho, S. (2008). Self-efficacy, metacognition, and performance. *North American Journal of Psychology*, *10*(1), 165.

Cui, Y., Kumar, R., Chaudhuri, S., Gweon, G., & Rosé, C. P. (2009). Helping agents in VMT. In *Studying virtual math teams* (pp. 335-354). Springer US.

Danker, J. F., & Anderson, J. R. (2007). The roles of prefrontal and posterior parietal cortex in algebra problem solving: A case of using cognitive modeling to inform neuroimaging data. *Neuroimage*, *35*(3), 1365-1377.

Dasgupta, S. (2011). Two faces of active learning. *Theoretical computer science*, *412*(19), 1767-1781.

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

Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport*, *6*(16), 2153-2157.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Cognitive Neuroscience, Journal of*, *8*(1), 47-68.

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

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*(2), 219-250.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in neurosciences*, *21*(8), 355-361.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive neuropsychology*, *20*(3-6), 487-506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*(5416), 970-974.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., & Mazoyer, B. (1996). Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia*, *34*(11), 1097-1106.

Desoete, A., Roeyers, H., & De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *Journal of learning disabilities*, *37*(1), 50-61.

Dew, K. H., Galassi, J. P., & Galassi, M. D. (1984). Math anxiety: Relation with situational test anxiety, performance, physiological arousal, and math avoidance behavior. *Journal of Counseling Psychology*, *31*(4), 580.

Diamond, D. M., Fleshner, M., Ingersoll, N., & Rose, G. (1996). Psychological stress impairs spatial working memory: relevance to electrophysiological studies of hippocampal function. *Behavioral neuroscience*, *110*(4), 661.

Druck, G., Settles, B., & McCallum, A. (2009, August). Active learning by labeling features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 81-90). Association for Computational Linguistics.

Duffy, Gerald G., and Laura R. Roehler. "Improving reading instruction through the use of responsive elaboration." *The Reading Teacher* (1987): 514-520.

Duncko, R., Johnson, L., Merikangas, K., & Grillon, C. (2009). Working memory performance after acute exposure to the cold pressor stress in healthy volunteers. *Neurobiology of learning and memory*, *91*(4), 377-381.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production?. *Psychology and aging*,*13*(4), 597.

Dunlosky, J., Hertzog, C., Kennedy, M. R., & Thiede, K. W. (2005). The Self-Monitoring Approach For Effective Learning. *Cognitive Technology*.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83-87.

Eccles, J. S. (1986). Gender-roles and women's achievement. *Educational Researcher*, 15-19.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological bulletin*, *136*(1), 103.

Elzinga, B. M., & Roelofs, K. (2005). Cortisol-induced impairments of working memory require acute sympathetic activation. *Behavioral neuroscience*, *119*(1), 98.

Markant, D., & Gureckis, T. M. (2010). Category learning through active sampling. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 248-253).

Erickson, S., & Heit, E. (2013). Math and metacognition: Resolving the paradox. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2255-2260).

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409-434.

Faust, M. W. (1996). Mathematics anxiety effects in simple and complex addition. *Mathematical Cognition*, *2*(1), 25-62.

Fayol, M., & Seron, X. (2005). About numerical representations: Insights from neuropsychological, experimental, and developmental studies.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive psychology*, *44*(1), 33-66.

Fennema, E. (1989). The study of affect and mathematics: A proposed generic model for research. In *Affect and mathematical problem solving* (pp. 205-219). Springer New York.

Ferry, T. R., Fouad, N. A., & Smith, P. L. (2000). The role of family context in a social cognitive model for career-related choice behavior: A math and science perspective. *Journal of Vocational Behavior*, *57*(3), 348-364.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research*, *65*(4), 250-259.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task?. *Psychological Research*, *65*(4), 242-249.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task?. *Psychological Research*, *65*(4), 242-249.

Fischer, M. H., & Fias, M. H. (2005). Spatial representation of numbers. In *The handbook of mathematical cognition*. Psychology Press New York.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819-824.

Furner, J. M., & Berman, B. T. (2003). Review of Research: Math Anxiety: Overcoming a Major Obstacle to the Improvement of Student Math Performance. *Childhood education*, *79*(3), 170-174.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of experimental child psychology*, *49*(3), 363-383.

Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychological bulletin*, *114*(2), 345.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction.*American Psychologist*, *50*(1), 24.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental psychology*, *27*(3), 398.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of experimental child psychology*,*77*(3), 236-263. Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of experimental child psychology*, *88*(2), 121-151.

Geist, E. (2010). The anti-anxiety curriculum: Combating math anxiety in the classroom. *Journal of Instructional Psychology*, *37*(1), 24.

Gelman, R., and C. R. Gallistel. "The child's understanding of number." *Harvard Univiversity Press, Cambridge* (1978).

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Archives of Neurology & Psychiatry*, *44*(2), 398-408.

Ginsburg, H. P., Klein, A., & Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice.

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). *Scientific reasoning across different domains* (pp. 345-371). Springer Berlin Heidelberg.

Göbel, S., Walsh, V., & Rushworth, M. F. (2001). The mental number line and the human angular gyrus. *Neuroimage*, *14*(6), 1278-1289.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496-499.

Groen, G., & Resnick, L. B. (1977). Can preschool children invent addition algorithms?. *Journal of Educational Psychology*, *69*(6), 645.

Gunnar, M. R., Connors, J., & Isensee, J. (1989). Lack of stability in neonatal adrenocortical reactivity because of rapid habituation of the adrenocortical response. *Developmental psychobiology*, *22*(3), 221-233.

Haith, M., & Benson, J. (1998). Infant cognition. In W. Damon (Series Ed.), D. Kuhn, & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, language, and perception* (5th ed., pp. 199-254). New York: Wiley.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology*, *93*(3), 615.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of educational research*, *66*(2), 99-136.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi–free–ranging rhesus monkeys. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*(1445), 829-833.

Heit, E., & Erickson, S. (2011). In praise of secular Bayesianism. *Behavioral and Brain Sciences*, *34*(04), 202-202.

Hellhammer, D. H., Hubert, W., & Schürmeyer, T. (1985). Changes in saliva testosterone after psychological stimulation in men. *Psychoneuroendocrinology*, *10*(1), 77-81.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163-171.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for research in mathematics education*, 33-46.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, *82*(3), 375-386.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of educational research*, *67*(1), 88-140.

Hopko, D. R., Ashcraft, M. H., Gute, J., Ruggiero, K. J., & Lewis, C. (1998). Mathematics anxiety and working memory: Support for the existence of a deficient inhibition mechanism. *Journal of Anxiety Disorders*, *12*(4), 343-355.

Hopko, D. R., Crittendon, J. A., Grant, E., & Wilson, S. A. (2005). The impact of anxiety on performance IQ. *Anxiety, Stress & Coping*, *18*(1), 17-35.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*,*6*(6), 435-448.

Hubert, W., & de Jong-Meyer, R. (1989). Emotional stress and saliva cortisol response. *Journal of clinical chemistry and clinical biochemistry. Zeitschrift für klinische Chemie und klinische Biochemie*, *27*(4), 235.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, *123*(3), 284.

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *SCIENCE-NEW YORK THEN WASHINGTON-*, *314*(5799), 599.

Jaccard, J., Dodge, T., & Guilamo-Ramos, V. (2005). Metacognition, risk behavior, and risk outcomes: the role of perceived intelligence and perceived knowledge. *Health Psychology*, *24*(2), 161.

Jackson, C. D., & Leffingwell, R. J. (1999). The role of instructors in creating math anxiety in students from kindergarten through college. *The Mathematics Teacher*, 583-586.

Jakobsson, N. (2012). Gender and confidence: are women underconfident?. *Applied Economics Letters*, *19*(11), 1057-1059.

Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presented fields of dots. *The Journal of Psychology*,*30*(2), 363-392.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(04), 169-188.

Jones, K. V., Copolov, D. L., & Outch, K. H. (1986). Type A, test performance and salivary cortisol. *Journal of psychosomatic research*, *30*(6), 699-707.

Jordan, N. C., & Montani, T. O. (1997). Cognitive Arithmetic and Problem Solving A Comparison of Children with Specific and General Mathematics Difficulties. *Journal of learning disabilities*, *30*(6), 624-634.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child development*, *74*(3), 834-850.

Junker, B. W., Koedinger, K. R., & Trottini, M. (2000, July). Finding improvements in student models for intelligent tutoring systems via variable selection for a linear logistic test model. In *Annual North American Meeting of the Psychometric Society*.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175-219.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American journal of psychology*, 498-525.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1), 92-107.

Kiefer, M., & Dehaene, S. (1997). The time course of parietal activation in single-digit multiplication: Evidence from event-related potentials. *Mathematical cognition*, *3*(1), 1-30.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*(5), 825-832.

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology*, *22*(3), 150-169.

Kitchner, K. S. (1983). Cognition, metacognition, and epistemic cognition. *Human development*, *26*(4), 222-232.

Kloosterman, P. (1988). Self-confidence and motivation in mathematics. *Journal of Educational Psychology*, *80*(3), 345.

Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. In *Animal mind—human mind* (pp. 201-224). Springer Berlin Heidelberg.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(2-3), 211-224.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*(4), 511-550.

Koedinger, K. R. & Anderson, J. R. (1993). Effective use of intelligent software in high school math classrooms. *In Proceedings of the World Conference on Artificial Intelligence in Education*, (pp. 241-248).Charlottesville, VA: Association for the Advancement of Computing in Education.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *Education (IJAIED)*, *8*, 30-43.

Koedinger, K. R., Corbett, A. T., Ritter, S., & Shapiro, L. (2000). Carnegie Learning's Cognitive Tutor: Summary Research Results. White paper. Available from Carnegie Learning Inc., 1200 Penn Avenue, Suite 150, Pittsburgh, PA 15222.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219-224.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281-310.

Krueger, L. E. (1982). Single judgments of numerosity. *Perception & Psychophysics*, *31*(2), 175-182.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of personality and social psychology*, *82*(2), 180.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science*, *9*(5), 178-181.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of cognition and Development*, *1*(1), 113-129.

Landsmann, L. T., & Karmiloff-Smith, A. (1992). Children's understanding of notations as domains of knowledge versus referential-communicative tools.*Cognitive development*, *7*(3), 287-300.

Langdon, D. W., & Warrington, E. K. (1997). The abstraction of numerical relations: a role for the right hemisphere in arithmetic?. *Journal of the International Neuropsychological Society*, *3*(03), 260-268.

Larson, M. C., Gunnar, M. R., & Hertsgaard, L. (1991). The effects of morning naps, car trips, and maternal separation on adrenocortical activity in human infants. *Child Development*, *62*(2), 362-372.

Le Clec'H, G., Dehaene, S., Cohen, L., Mehler, J., Dupoux, E., Poline, J. B., & Le Bihan, D. (2000). Distinct cortical areas for names of numbers and body parts independent of language and input modality. *Neuroimage*, *12*(4), 381-391.

Lee, K. M., & Kang, S. Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, *83*(3), B63-B68.

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: a functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Annals of Neurology*, *48*(4), 657-661.

Lewis, M., & Thomas, D. (1990). Cortisol release in infants in response to inoculation. *Child Development*, *61*(1), 50-59.

Lodico, M. G., Ghatala, E. S., Levin, J. R., Pressley, M., & Bell, J. A. (1983). The effects of strategy—Monitoring training on children's selection of effective memory strategies. *Journal of Experimental Child Psychology*,*35*(2), 263-277.

Lupien, S. J., Gillin, C. J., & Hauger, R. L. (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A dose–response study in humans. *Behavioral neuroscience*, *113*(3), 420.

Lyons, I. M., & Beilock, S. L. (2012). Mathematics anxiety: separating the math from the anxiety. *Cerebral Cortex*, *22*(9), 2102-2110.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 520-540.

MacLullich, A. M., Edmond, C. L., Ferguson, K. J., Wardlaw, J. M., Starr, J. M., Seckl, J. R., & Deary, I. J. (2004). Size of the neocerebellar vermis is associated with cognition in healthy elderly men. *Brain and cognition*, *56*(3), 344-348.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,*10*(4), 663.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1.

Markant, D., & Gureckis, T. M. (2010). Category learning through active sampling. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 248-253).

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of Multidimensional Self-Concept and Core Personality Constructs: Construct Validation and Relations to Well-Being and Achievement. *Journal of personality*, *74*(2), 403-456.

Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., & Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math-anxiety. *Emotion*, *11*(4), 1000.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1), 107-157.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of educational psychology*, *82*(1), 60.

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage*, *12*(4), 357-365.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1084.

Metcalfe, J. E., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. The MIT Press.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning?. *Journal of Experimental Psychology: General*, *131*(3), 349.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163.

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal*, *34*(2), 365-394.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: reinterpreting overconfidence in low-performing students. *Journal of experimental psychology: learning, memory, and cognition*, *37*(2), 502.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (Vol. 27).

Mizoguchi, K., Yuzurihara, M., Ishige, A., Sasaki, H., Chui, D. H., & Tabira, T. (2000). Chronic stress induces impairment of spatial working memory because of prefrontal dopaminergic dysfunction. *The Journal of Neuroscience*, *20*(4), 1568-1574.

Morgan, P., & Ritter, S. (2002). An experimental study of the effects of Cognitive Tutor® Algebra I on student knowledge and attitude. *Pittsburgh, PA: Carnegie Learning, Inc*.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature:* 216 (159-160).

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*,*79*(1), 1-37.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, *2*(4), 267-270.

Newcombe, N. S. (2002). The nativist-empiricist controversy in the context of recent research on spatial and quantitative development. *Psychological Science*, *13*(5), 395-401.

Nicholls, J. G., & Miller, A. T. (1984). Reasoning about the ability of self and others: A developmental study. *Child development*, 1990-1999.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological review*, *84*(3), 231.

Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*,*82*(1), B25-B33.

O'Brien, J. T., Lloyd, A., McKeith, I., Gholkar, A., & Ferrier, N. (2004). A longitudinal study of hippocampal volume, cortisol levels, and cognition in older depressed subjects. *American Journal of Psychiatry*, *161*(11), 2081-2090.

Oei, N. Y. L., Everaerd, W. T. A. M., Elzinga, B. M., Van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: an association with cortisol levels and memory retrieval. *Stress*, *9*(3), 133-141.

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, *8*(2), 103-179.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement?. *Contemporary Educational Psychology*, *26*(3), 291-310.

Pajares, F., & Valiante, G. (2002). STUDENTS'SELF-EFFICACY IN THEIR SELF-REGULATED LEARNING STRATEGIES: A DEVELOPMENTAL PERSPECTIVE. *Psychologia*, *45*(4), 211-221.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253-282.

Paris, S. G., & Byrnes, J. P. (1989). The constructivist approach to self-regulation and learning in the classroom. In *Self-regulated learning and academic achievement* (pp. 169-200). Springer New York.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational psychologist*, *36*(2), 89-101.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. *Dimensions of thinking and cognitive instruction*, *1*, 15-51.

Pesenti, M., Thioux, M., Seron, X., & Volder, A. D. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Cognitive Neuroscience, Journal of*, *12*(3), 461-479.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage*, *15*(2), 435-446.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage*, *14*(5), 1013-1026.

Plake, B. S., & Parker, C. S. (1982). The development and validation of a revised version of the mathematics anxiety rating scale. *Educational and Psychological Measurement*, *42*(2), 551-557.

Pressley, M. (2002). Effective beginning reading instruction. *Journal of Literacy Research*, *34*(2), 165-188.

Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 270-288.

Pugalee, D. K. (2001). Writing, mathematics, and metacognition: Looking for connections through students' work in mathematical problem solving. *School Science and Mathematics*, *101*(5), 236-245.

Qin, Y., Carter, C. S., Silk, E. M., Stenger, V. A., Fissell, K., Goode, A., & Anderson, J. R. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(15), 5686-5691.

Qin, S., Hermans, E. J., van Marle, H. J., Luo, J., & Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological psychiatry*, *66*(1), 25-32.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*(2), 187-202.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2p1), 274.

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, *21*(4), 387-399.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *The Quarterly Journal of Experimental Psychology: Section A*,*55*(4), 1127-1139.

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of counseling Psychology*, *19*(6), 551.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review*, *14*(2), 249-255.

Rittle-Johnson, B., & Koedinger, K. R. (2002). Comparing Instructional Strategies for Integrating Conceptual and Procedural Knowledge.

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing knowledge scaffolds to support mathematical problem solving. *Cognition and Instruction*, *23*(3), 313-349.

Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, *93*(2), 346.

Roland, P. E., & Friberg, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, *53*(5), 1219-1243.

Rosselli, M., & Ardila, A. (1989). Calculation deficits in patients with right and left hemisphere damage. *Neuropsychologia*, *27*(5), 607-617.

Rourke, B. P., & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning perspectives from neurology and neuropsychology. *Journal of Learning disabilities*, *30*(1), 34-46.

Rueckert, L., Lange, N., Partiot, A., Appollonio, I., Litvan, I., Le Bihan, D., & Grafman, J. (1996). Visualizing cortical activation during mental calculation with functional MRI. *Neuroimage*, *3*(2), 97-103.

Scher, D. P. (2002). *Students' conceptions of geometry in a dynamic geometry software environment* (pp. 1-147).

Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: An integrative review. *Metacognition, cognition, and human performance*, *1*, 57-109.

Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive science*, *7*(4), 329-363.

Schoenfeld, A. H. (1987). What's All the Fuss About Metacognitlon. *Cognitive science and mathematics education*, *189*.

Schraw, G. (1996). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *The Journal of Experimental Education*, *65*(2), 135-146.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*(1-2), 111-139.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, *96*(4), 523.

Seo, K-H, & Ginsburg, H. P. (2003). "You've got to carefully read the math sentence.. .": Classroom context and children's interpretations of the equals sign. *In A. J. Baroody & A. Dowker (Eds.), The development of arithmetic concepts and skills.* Mahwah, NJ: Erlbaum.

Seron, X., Pesenti, M., Noël, M. P., Deloche, G., & Cornet, J. A. (1992). Images of numbers, or "When 98 is upper left and 6 sky blue". *Cognition*,*44*(1), 159-196.

Druck, G., Settles, B., & McCallum, A. (2009, August). Active learning by labeling features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 81-90). Association for Computational Linguistics.

Seymour, S. E., Reuter-Lorenz, P. A., & Gazzaniga, M. S. (1994). The disconnection syndrome. *Brain*, *117*(1), 105-115.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.

Sharon, T., & Wynn, K. (1998). Individuation of actions from continuous motion. *Psychological Science, 9*(5), 357-362.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and cognition, 9*(2), 313-323.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405-410.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*(3), 258.

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

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. *Developing cognitive competence: New approaches to process modeling*, 31-76.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A "non-numerical" account. *Cognitive Development, 12*(3), 349-372.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204.

Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science, 30*(4), 759-774.

Son, L. K., & Sethi, R. (2009). Adaptive learning and the allocation of time. *Adaptive Behavior*.

Spada, M. M., Nikčević, A. V., Moneta, G. B., & Wells, A. (2008). Metacognition, perceived stress, and negative emotion. *Personality and Individual Differences, 44*(5), 1172-1181.

Spangler, G., & Scheubeck, R. (1993). Behavioral organization in newborns and its relation to adrenocortical and cardiac activity. *Child Development,64*(2), 622-633.

Spelke, E., & Dehaene, S. (1999). Biological foundations of numerical thinking. *Trends in Cognitive Sciences, 3*(10), 365-366.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*(1), 45-88.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology,35*(1), 4-28.

Sperling, R.A., Howard, B.C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10(2)*, 117-139.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge* (pp. 451-473). Cambridge, MA: Mit Press.

Stahl, G., Rosé, C. P., O'Hara, K., & Powell, A. B. (2010, July). Supporting group math cognition in virtual math teams with software conversational agents. In *First North American GeoGebra Conference* (p. 196).

Stahl, G., Rosé, C. P., & Goggins, S. (2011). Analyzing the discourse of geogebra collaborations. *Essays in Computer-Supported Collaborative Learning, 186*, 33-40.

Stanescu-Cosson, R., Pinel, P., van de Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia. *Brain*, *123*(11), 2240-2255.

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033-1035.

Starkey, P. (1992). The early development of numerical reasoning.*Cognition*, *43*(2), 93-126.

Starkman, M. N., Giordani, B., Berent, S., Schork, M. A., & Schteingart, D. E. (2001). Elevated cortisol levels in Cushing's disease are associated with cognitive decrements. *Psychosomatic Medicine*, *63*(6), 985-993.

Starkman, M. N., Giordani, B., Gebarski, S. S., Berent, S., Schork, M. A., & Schteingart, D. E. (1999). Decrease in cortisol reverses human hippocampal atrophy following treatment of Cushing's disease. *Biological psychiatry*,*46*(12), 1595-1602.

Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 1146-1152.

Stuart, V. B. (2000). Math curse or math anxiety?. *Teaching children mathematics*, *6*(5), 330.

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, *8*(2), 16.

Temple, C. M., & Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(3), 733-752.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of experimental psychology: Learning, Memory, and Cognition*, *25*(4), 1024.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology*, *95*(1), 66.

Thioux, M., Pesenti, M., Costes, N., De Volder, A., & Seron, X. (2005). Task-independent semantic activation for numbers and animals. *Cognitive Brain Research*, *24*(2), 284-290.

Tiemensma, J., Kokshoorn, N. E., Biermasz, N. R., Keijser, B. J. S., Wassenaar, M. J., Middelkoop, H. A., & Romijn, J. A. (2010). Subtle cognitive impairments in patients with long-term cure of Cushing's disease.*The Journal of Clinical Endocrinology & Metabolism*, *95*(6), 2699-2714.

Tong, S., & Koller, D. (2001, August). Active learning for structure in Bayesian networks. In *International joint conference on artificial intelligence* (Vol. 17, No. 1, pp. 863-869). LAWRENCE ERLBAUM ASSOCIATES LTD.

Townsend, C., & Heit, E. (2010). Metacognitive judgments of improvement are uncorrelated with learning rate. In *Proceedings of the 32nd annual conference of the cognitive science society. Austin, TX: Cognitive Science Society*.

Townsend, C. L., & Heit, E. (2011). Judgments of learning and improvement. *Memory & cognition*, *39*(2), 204-216.

Tucker, A. B., Kelemen, C. F., & Bruce, K. B. (2001, February). Our curriculum has become math-phobic!. In *ACM Sigcse Bulletin* (Vol. 33, No. 1, pp. 243-247). ACM.

van Eck, Marleen, et al. "The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol." *Psychosomatic medicine* 58.5 (1996): 447-458.

Van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*(2), 163-170.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological review*, *97*(4), 488.

Veenman, M. V., & Verheij, J. (2001). Technical students' metacognitive skills: Relating general vs. specific metacognitive skills to study success. *Learning and Individual differences*, *13*(3), 259-272.

Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and individual differences*, *15*(2), 159-176.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning*, *1*(1), 3-14.

Wender, K. F., & Rothkegel, R. (2000). Subitizing and its subprocesses. *Psychological Research*, *64*(2), 81-92.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 135.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130-137.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571-580.

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, *80*(2), 210.

Wolters, C. A. (1999). The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance. *Learning and individual differences*, *11*(3), 281-299.

Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of personality and social psychology*, *56*(3), 407.

Wood, A. M., Boyce, C. J., Moore, S. C., & Brown, G. D. (2012). An evolutionary based social rank explanation of why low income predicts mental distress: A 17 year cohort study of 30,000 people. *Journal of affective disorders*, *136*(3), 882-888.

Wood, A. M., Brown, G. D., & Maltby, J. (2012). Social norm influences on evaluations of the risks associated with alcohol consumption: Applying the rank-based decision by sampling model to health judgments. *Alcohol and alcoholism*, *47*(1), 57-62.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749-750.

Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in cognitive sciences*, *2*(8), 296-303.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.

Young, C. B., Wu, S. S., & Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychological Science*, 0956797611429134.

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *Neuroimage*, *13*(2), 314-327.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of educational Psychology*, *82*(1), 51.