Do College Students Learn to Critically Evaluate Claims?

A Cross-Sectional Study of Freshmen and Senior Psychology Majors

by

Fernando Rodriguez

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education and Psychology)
in The University of Michigan
2011

Doctoral Committee:

Associate Professor Priti R. Shah, Chair
Professor Patricia M. King
Professor Kevin F. Miller
Research Professor Stuart A. Karabenick

Dedication

To mom, dad, and sis

Acknowledgements

I am grateful to a number of people who have guided my development throughout my time here at Michigan. Dr. Priti Shah, my dissertation committee chair, has been a wonderful mentor. Her enthusiastic commitment to my research and her intellectual insights has given me a strong theoretical foundation to pursue further research on students' critical thinking development. I look forward to continuing our work.

Dr. Kevin Miller has been a great advisor. His observations, insights, and suggestions for improving my dissertation were invaluable.

Dr. Stuart Karabenick provided valuable advice regarding my methodological approaches. He truly helped improve the design of my dissertation studies.

Dr. Patricia King helped me clarify many theoretical issues I had regarding the relationship between personal epistemology and thinking. I appreciated her insightful comments, which tremendously helped improve the overall quality of my dissertation.

I would like to acknowledge the Horace H. Rackham School of Graduate Studies for funding my dissertation project. I would also like to thank the Gayle Morris Sweetland Center for Writing who provided me funding and office space during the final months of my dissertation writing. The Evolutionary and Ontogenetic Dynamics (LIFE) graduate program provided me with invaluable research training well as funding for attending their academies. I would like to thank the Combined Program in Education & Psychology for my research training and for funding me throughout my graduate career.

I owe a big thanks to my research assistants, Annalyn Ng, Sam Gross, and Diego Rizzo, who helped in every aspect of the research process. I could not have collected all of my research data without their help and commitment. I would also like to thank the students who participated in my research studies.

I would like to thank all my friends in Ann Arbor who made my time in graduate school memorable. My parents, Fernando and Armida Rodriguez, and my sister, Elizabeth Rodriguez, have supported me the most and are my greatest inspiration. Finally, my wife Annie Ro has been my emotional center and my very best friend. I could not have done it without her.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Although prior research has used a variety of methods for studying college students' critical thinking skills, few have provided an observational account of the reasoning strategies students develop while in college. The goal of my dissertation was to observe how college students utilized experiential and scientific reasoning strategies when evaluating evidence-based claims. I considered these questions in light of the dual process model of reasoning, which views reasoning in terms of quick and intuitive experiential systems or effortful and deliberate rational systems. I also examined the relationship between students' scientific knowledge, thinking dispositions, and prior beliefs on their experiential and scientific reasoning outcomes. Study 1 asked underclassmen to evaluate experimental studies. When not explicitly asked to think critically, students used experiential strategies over scientific strategies to evaluate evidence. Study 2 compared college freshmen with senior psychology majors and additionally examined how alluring anecdotal stories influenced the reasoning process. Students were asked to evaluate a set of flawed evidence-based claims. Half received an alluring anecdotal story alongside each claim. Students agreed more and provided more experiential evaluations when the claims contained alluring anecdotal stories. Seniors were better at evaluating the claims scientifically and provided more in-depth evaluations when compared to freshmen. Differences in students' scientific knowledge and open-minded thinking were related to providing more scientific evaluations. Students' prior beliefs showed the opposite relationship; when claims fit students' prior beliefs, they agreed more with the claims and provided more experiential evaluations. My results show that college training in psychology may teach students how to evaluate claims scientifically, but alluring information and belief-basis reasoning remain key barriers towards critical thinking. However, open-mindedness may be an important characteristic that helps promote the ability to reason independently from one's prior beliefs.

CHAPTER I

Introduction

We encounter a number of claims in our day-to-day lives that influence our views and decisions, such as whether video games are bad for children or if certain foods prevent disease. For instance, it is very common for news articles to describe the latest scientific studies and discuss their implications. Although scientific research studies are more valid than other sources of evidence, it is still important to be critical about how news articles present and interpret scientific studies. Articles can make flawed assertions about scientific results. They may over-interpret studies containing weak findings as being critically important, or they may state that a causal relationship exists from correlational findings. Additionally, scientific news reports often include vivid anecdotal stories, which may influence people to believe that the article is more valid that what the evidence implies. Being able to evaluate claims is not only relevant in the context of evaluating scientific news reports but also when listening to speakers or political leaders, whose goals are to persuade people to accept their viewpoints. Additionally, advertisers also use 'scientific data' and anecdotal testimonials to convince people to buy their products.

There is a danger in accepting information at face value or in making decisions based on whether something seems intuitively appealing. Reasoning this way can lead to adopting unfounded beliefs, making poor decisions, and diminishing one's ability to make accurate judgments about the world. People need the important critical thinking skills to make good choices in their day-to-day lives and college may be the most important place where students learn to do this. Understanding whether college helps students develop these skills is therefore worthwhile.

The goal of my dissertation is to understand how college students develop the ability to think critically about everyday claims. I pay special attention to evidence-based claims, like the kind people come across when reading news reports about the latest social-scientific studies. The ability to evaluate these types of claims requires scientific reasoning skills. This strategy for reasoning involves assessing how a research study was designed, how observations were measured and quantified, whether important factors were controlled for, and whether flawed interpretations were made from the findings, to name a few (Schunn & Anderson, 1999; Zimmerman, 2000). More fundamentally, however, the purpose of reasoning scientifically about evidence-based claims is premised on the idea that in order conclude whether a claim or argument is genuinely valid one has to examine the quality of the supporting evidence, a process is known as theory-evidence coordination (Kuhn, 1992).

Educators want students to make well-informed judgments when presented with knowledge claims, assertions, or arguments in their everyday lives. However, a recent study by Arum and Roska (2011) found that students' critical thinking skills improved very little during the first two years in college. Other scholars have also questioned whether college graduates learn important scientific reasoning skills (Bullock, Sodian, & Koerber, 2008; Kuhn, 2009; Sodian & Bullock, 2008). Despite these concerns, the extent to which students are learning to think scientifically about claims has yet to be fully understood.

One reason is because some disciplines emphasize scientific thinking and reasoning more than others. Although one of the goals of college is teach every student to think more critically about claims and arguments, studies have found that these abilities develop differently across disciplines. For instance, students majoring in the social sciences, who learn about the scientific logic behind conducting and evaluating experiments, have been shown to be better at evaluating experimental and statistical evidence than students majoring in math, who do better evaluating the validity of logical statements. Therefore, the extent to which we can understand whether college students learn to think critically is limited to the set of students we observe and the field of study we ask them to think critically about. Because of this, my dissertation only considers how college students evaluate evidence-based claims and whether seniors majoring in psychology are better able to evaluate these claims more scientifically than freshmen students.

Although numerous studies have examined college students' critical thinking development, few have offered an observational account of the reasoning strategies students use to evaluate claims and arguments. Traditional methods for assessing critical thinking have focused on using multiple-choice items, asking students to evaluate information using Likert-type scales, or scoring students' responses in terms of whether they reflect good or poor critical thinking. These methods are useful for understanding whether students are thinking critically, but an observational approach is also needed that captures how students' reasoning processes change as a function of college.

Dual process models provide a useful framework for observing how students evaluate evidence. This model describes reasoning as stemming from two basic cognitive systems: the autonomous and the non-autonomous (Evans, 2003; Stanovich, 1999; Stanovich & West, 2000). Where the former relies on quick and intuitive processing the other is more deliberate, effortful, and controlled. Studies have generally found that people prefer using their intuition, belief, and experiences when reasoning, which oftentimes lead to making poor inferences from information and evidence (Evans, 2007; Klaczynski, 2001a; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Robinson, 2000). Using this dual process approach, I examine if psychology majors, who are taught important principles of scientific reasoning, use less intuitive-experiential forms of reasoning and more rational-analytic approaches – such as scientific reasoning – when evaluating evidence-based claims.

Viewing critical thinking as the ability to reason scientifically only tells half of the story, however. Critical thinking also requires reflective skills, such as the ability to notice when it is appropriate to think critically and to separate one's beliefs and experiences when evaluating evidence. For instance, belief-bias reasoning has been shown to undermine how well individuals think critically about evidence. That is, individuals are more likely to evaluate evidence favorably when it confirms their previous beliefs and less favorably when evidence is belief-threatening.

Several studies have demonstrated that intellectual ability, as indexed by ACT scores, verbal intelligence, and fluid intelligence measures, is unrelated to avoiding belief-basis reasoning (Macpherson & Stanovich, 2007; Stanovich, 1999, 2009; Stanovich & West, 2008). Instead, important dispositional factors have been linked to differences in individuals' tendencies to engage in belief-bias reasoning (Halpern, 1999; Stanovich & West, 1998). Some students are more motivated to engage in effortful thinking than others. Additionally, students also differ by their willingness to consider different viewpoints and ideas. Finally, students have different epistemic beliefs about what constitutes knowledge and how knowledge is justified (Hofer & Pintrich, 1997; King & Kitchener, 1994; King, Wood, & Mines, 1990). These dispositions are important to consider when observing students' reasoning processes, since they help explain why some students' reasoning is more effortful and less biased than others.

Another issue I consider is how alluring anecdotal stories influence how college students perceive and evaluate evidence-based claims. It is very common for scientific news articles to include short anecdotal stories prior to describing a research study. Although anecdotal stories help readers understand how research studies relate to real events and experiences, such stories may also elicit experiential processing in which individuals evaluate articles on based on how well stories fits with their own personal experiences or whether the story seems believable (Dahlstrom, 2010; Strange & Leung, 1999; Winterbottom, Bekker, Conner, & Mooney, 2008). Although the ability to resist alluring information is considered an important critical thinking skill, the extent to which students can avoid being influenced by anecdotal information when evaluating evidence-based claims has not been thoroughly studied.

The purpose of my dissertation is to shed light on whether students learn to think critically about evidence-based claims. Given that social science majors are taught to think scientifically, I examine whether seniors majoring in psychology are better able to reason scientifically when evaluating evidence-based claims than freshman students. I additionally want to understand whether college potentially changes students thinking dispositions and epistemic beliefs, since it has implications for understanding whether college training also helps students' avoid belief-bias reasoning. Finally, I also examine whether college training helps students resist the allure of persuasive information. In the literature review that follows, I define critical thinking, overview key studies, discuss the different lines of research, and propose a model for understanding college students' critical thinking development. Afterwards, I present two studies in which I observe how students evaluate evidence while also exploring how differences in students' evaluations are related to their thinking dispositions and class standing.

Study 1 was deigned to provide an observational account of students' reasoning strategies. For this study, I asked college underclassmen to evaluate a set of research studies, once in an informal context and again in a critical thinking context and compared how their reasoning strategies differed in these two situations. I also considered how individual differences in students' need for cognition and open-mindedness was related to the type of reasoned evaluation students provided. In Study 2a, I compared freshmen students with seniors majoring in psychology. I wanted to understand whether seniors used more scientific reasoning strategies than freshmen. I also compared whether seniors reported being more open-minded, more rationally (versus experientially) oriented, and more reflective in their epistemic beliefs about the nature and sources of knowledge. Finally, I examined whether seniors were less persuaded by anecdotal information than freshmen. In Study 2b, I address alternative explanations regarding the results of Study 2a. Lastly, I discuss how my results address some gaps in the literature and advance how scholars understand the nature of college students' critical thinking development.

CHAPTER II

Review of the Literature

Understanding what it means to think critically

There has been much debate about how to best define and measure critical thinking, since it occurs in a wide range of tasks and involves a wide array of skills and abilities (Ennis, 1989; McPeck, 1981; Williams, 1999). Thus, attempts to provide overarching definitions of critical thinking have been criticized for being vague while definitions that itemize critical thinking into specific skills and strategies do not offer clear conceptual explanations. However, discussions, most notably by McPeck (1981) and Williams (1999), have helped clarify how we understand critical thinking in three important ways. These clarifying points serve as my guiding model for how I define critical thinking in the domain of evaluating claims and arguments.

The first involves how critical thinking is conceptualized. Generally speaking, critical thinking is a process of thinking *deliberately* and *reflectively*. This deliberate component involves using specific reasoning skills and strategies when deciding what to believe or do (Bensley & Haynes, 1995; Halpern, 1998). The reflective component, on the other hand, involves metacognitive processes, such as understanding how beliefs, biases, dispositions, and abilities influence the critical thinking process (Giancarlo & Facione, 2001; King, 2000; Kuhn, 1999, 2000).

The second clarification is that critical thinking should be explicitly defined and measured according to the task that is being considered (McPeck, 1981; Williams, 1999). People think critically in a variety of situations, such as when making decisions, developing arguments, and forming theories, to name a few. What it means to think critically will therefore function differently across these different tasks. For instance, making managerial decisions, where one has to choose a single course of action from several competing options, involves generating hypotheses about which action will lead to the most optimal outcome while also reflecting about whether one's decision will adversely affect some other outcome, like social cohesion among team members. Evaluating claims, on the other hand, involves assessing whether the rationale for the claim is grounded in evidence, and if so, whether the evidence is strong and convincing enough to support the claim. This also involves reflective skills, like monitoring whether one's personal beliefs are biasing how a claim is being evaluated.

The third and final point is that critical thinking is dependent on knowledge. Having sufficient knowledge in a given domain provides individuals with the necessary understanding and relevant skills that are needed to reach well-reasoned conclusions (Leshowitz & Okun, 2011; Norcross, Gerrity, & Hogan, 1993; Schunn & Anderson, 1999). There has been some debate, however, about whether critical thinking skills are subject-specific. Some have argued that subject-specific knowledge is only useful in its given domain (McPeck, 1981), whereas others posits that subject-specific knowledge can aid critical thinking across other domains (Ennis, 1989; Smith, 2002). In the context of scientific reasoning, work by Schunn and Anderson (1999) helped clarify part of this debate by differentiating between scientific reasoning skills that are domain-specific and domain-general. Domain-specific knowledge refers to understanding specific concepts, principles, and epistemological assumptions that govern a given domain. Domain-general knowledge, on the other hand, refers to the process of applying conceptual and procedural knowledge to problems and tasks. Although it is not entirely clear how domain-specific knowledge functions in other domains, studies have demonstrated that domain-general knowledge can be used to think and solve problems in other domains (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Kosonen & Winne, 1995; Nisbett, Fong, Lehman, & Cheng, 1987).

Critical thinking in the context of evaluating claims and arguments

For dissertation, I pay specific attention to understanding how college students critically evaluate claims and arguments. Under this context, I define critical thinking as a process of coordinating between theory and evidence, where the validity of a claim or an argument is critically evaluated by examining the strength and sources of the supporting evidence (Bensley, Crowe, Bernhardt, Buckner, & Allman, 2010; Kuhn, 1992; Stanovich & Stanovich, 2010). As an example, take the following claim regarding siblings and creativity,

> Younger siblings are more creative. As one parent notes, "Anyone who's around children will tell you that younger siblings are very unique. My youngest son, Jacob, for instance, spends his free time creating costumes and speaking in imaginary languages, while my oldest child, Brendon, prefers to watch TV and play videogames." Interestingly, a recent study also found that younger siblings scored 2 points higher on a creative test than older siblings.

Scholars note that the ability to coordinate between theory and evidence requires scientific reasoning skills (Koslowski, 1996; Kuhn, Iordanou, Pease, & Wirkala, 2008;

Zimmerman, 2000). These reasoning skills can be understood as involving domain-specific and domain-general knowledge. Domain-specific scientific knowledge refers to and individuals' scientific understanding, such as the idea that knowledge about the physical and social world is obtained by creating and evaluating experiments that test theories and ideas, and that as more evidence becomes available, we can know with greater confidence that a particular theory provides a valid explanation. Domain-general scientific knowledge, however, refers to the process of applying scientific concepts and procedures, like how to tests theories and structure experimental designs, to thinking about problems and tasks. In the context of evaluating claims, this would involve things like identifying evidence from other types of information, examining how evidence is constructed and measured, considering the role of other factors, and identifying strengths and flaws in the evidence. In the above claim, for example, one would immediately question whether a parent's anecdotal observations is a reliable source of evidence and whether a 2-point difference in a creativity score is meaningful enough to support the claim that younger siblings are more creative.

There have been a number of studies that have examined students' critical thinking skills in the context of evaluating claims and arguments. However, many of these studies have traditionally measured students' critical thinking using problems with pre-determined solutions, in which students are asked to choose the best response from a set of options (Lehman & Nisbett, 1990; Pascarella, Bohr, Nora, & Terenzini, 1996; Terenzini, Springer, Pascarella, & Nora, 1995). Other studies have also relied on using Likert-type scales to reflect whether students are thinking more or less critically (Bastardi, Uhlmann, & Ross, in press; Cacioppo, Petty, & Morris, 1983; Macpherson &

Stanovich, 2007; Stanovich & West, 2007). Although these methods provide us with a general index of students' critical thinking skills, they do little to tell us how students are reasoning when thinking critically. Accordingly, there have been recent calls to specify the specific scientific reasoning strategies students utilize when evaluating evidence (Kuhn et al., 2008; Schunn & Anderson, 1999; Sodian & Bullock, 2008).

Observing how students reason scientifically advances how we understand critical thinking processes in two important ways. First, it provides us with a clear picture of how students reason when evaluating evidence. Second, examining the patterns of students responses – whether they provide more experiential versus scientifically reasoned evaluations, whether they engage in belief-bias reasoning, and whether their scientific reasoning reflects surface versus deep level processing – provides valuable insights into understanding why some students are better able to think critically about evidence than others.

A major ongoing debate has been whether American college students are developing important critical thinking skills, especially those that are needed to evaluate claims and arguments. Arum and Roskas' (2011) work gained recent attention for showing that students' critical thinking abilities does improved little over the course of two years in college. Arum and Roska have also suggested that college students' critical thinking abilities have been steadily declining within the last decade. However, critical thinking has been defined and measured quite differently over the years and to varying degrees of success.

In the following sections I review some common ways critical thinking has been studied. In doing so, I try to distinguish between measures that reflect overarching skills versus task specific ones and point out some developments in measuring critical thinking. Using a dual process model for understanding reasoning, I argue that attending college helps students more away from using institutive-experiential reasoning systems and more towards rational-analytic systems when evaluating evidence. In examining this, I consider how domain-general scientific knowledge as well as reflective aspects of critical thinking, like thinking dispositions and epistemic beliefs, play a role in understanding how college students evaluate claims and arguments. Another issue I discuss is how the characteristics of a claim, like the presence of anecdotal stories, influences how students perceive and evaluate evidence. Finally, I point out some key issues in the research that have not been fully addressed and propose a model of critical thinking development that focuses on observing students' students' individual characteristics (e.g., scientific knowledge, thinking dispositions) and their reasoning strategies.

Measuring the development of college students critical thinking skills

Early work studied critical thinking under a problem solving framework, which primarily focused on how well students could correctly assess statements (Ennis, 1962). One of the earliest critical thinking measures, the Test of Critical Thinking, developed by Dressel & Mayhew (1954), assessed five abilities they believed to be essential for thinking critically. These were the ability to, 1) define a problem, 2) select pertinent information for the solution to a problem, 3) recognize stated and unstated assumptions, 4) formulate and select relevant and promising hypotheses, and 5) draw valid conclusions (Dressel & Mayhew, 1954, pgs. 179-180). The Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964) was another widely used measure during this time.

Borrowing it's ideas from the Test of Critical Thinking, The Watson-Glaser CTA assessed similar sets of skills, like the ability to draw inferences from information, recognizing assumptions, using deduction, drawing interpretations, and evaluating arguments. Using these measures, a number of longitudinal studies found that students made significant gains in critical thinking skills over the course of college (Dressel & Mayhew, 1954; Feldman & Newcomb, 1969; Lehmann, 1963; Rickert, 1967).

Since these were the first measures that attempted to capture critical thinking, they had several limitations. First, the items were not constructed to assess how well people thought critically in a given task, but rather they represented a wide variety of logical problems that were believed to capture some important aspect of critical thinking. For example, one of the inference problems on the Watson-Glaser CTA asked students to read a passage and provide one of five responses regarding the truthfulness of the stated facts: *true*, *partly true*, *insufficient data*, *partly false*, *false*. The following example was taken from McPeck (1981, p. 136).

> Two hundred eighth-grade students voluntarily attended a recent weekend student forum conference in a Midwestern city. At this conference, the topics of race relations and means of achieving lasting world peace were discussed, since these were the problems the students selected as being most vital in today's world.
>
> Stated Facts
>
> 1. As a group, the students who attend this conference showed a keener interest in humanitarian or broad social problems than have most eighth-grade students. (correct answer: *partly true*)
> 2. The majority of these students were between the ages of 17 and 18. (correct answer: *partly false*)
> 3. The students came from all sections of the country. (correct answer: *insufficient data*)
> 4. The students discussed only labor relations problems. (correct answer: *false*)
> 5. Some eight-grade students felt that discussion of race relations and means of achieving world peace might be worthwhile. (correct answer: *true*)

For these tasks, only one of the responses options was considered correct. Thus, getting the correct answer implied that one used important critical thinking skills to do so. Another limitation of these tests was that, as McPeck (1981) pointed out, the way the responses were scored did not reflect what participants were instructed to do. In the above example, participants were asked to base their responses using only the given information, but some correct scores required participants to use their 'general knowledge' to draw appropriate conclusions from the information.

Scholars of critical thinking measures have pointed out that it is important to differentiate between problems that are either well-structured or ill-structured (Churchman, 1971; King, 2000; Schraw, Dunkle, & Benedixon, 1995; Sternberg, 1982). Well-structured problems are self-contained, where all the relevant information is given and the premises are assumed to be true. Therefore, individuals can solve these problems with a high degree of certainty by applying logical rules and principles to reach the correct solution. In contrast, ill-structured problems are more complex in nature. They involve solving problems that people encounter in their everyday lives, like who to vote for, what to decide, and whether one should believe new scientific findings. These problems do not always have clear-cut solutions since people will have varying amounts of information and expertise. Therefore, under this second problem structure, thinking critically is not about the ability to reach a correct conclusion, but is instead the ability to reach a solution that is well reasoned and adequately justified.

Later work using this distinction led to a more nuanced understanding how college differentially impacted students' critical thinking development (Lehman, Lempert, & Nisbett, 1988; Lehman & Nisbett, 1990; Pascarella et al., 1996; Schraw et al., 1995). In Lehman and Nisbetts' (1990) longitudinal study, college students were followed from the start of their freshmen year to the end of their senior year. In order to examine whether students' disciplinary training was related to their ability to solve well-structured and ill-structured problems, Lehman and Nisbett sampled students from the natural sciences, humanities, social sciences, and psychology. Students were given a set of well-structured and ill-structured problems to solve and observed whether their responses improved over the course of college. For the well-structured problem, students solved a conditional-logic task in which they had to correctly select one of four cards based on a permission rule. For the ill-structured problem, students were given real world problems to think about (e.g., Why do rookies who perform well their first year have lower batting averages the following year?) and were assessed according to whether they selected a statistical explanation from a set of answers. As they predicted, natural science students showed the most improvement in the conditional-logic problem (65% gain) whereas students in social science and psychology showed no significant gains. Conversely, students in social science and psychology made the largest gains on the ill-structured problems (67% and 65%, respectively) compared to students in natural science and humanities.

Distinguishing between well-structured and ill-structured tasks has helped scholars understand and discuss critical thinking more precisely. Studies have also moved away from multiple choice and scale type measures towards measures that capture how well students thinking about everyday problems (Sandoval & Milwood, 2005; Toplak & Stanovich, 2003). For instance, in a study by Norcross and colleagues (1993) college students were asked to read a set of recommendations that were supported by research studies. These studies, however, had clear flaws in the methodology, which served to discredit the validity of the evidence. This included things such as lack of a comparison group or bias in the sampling. When asked to evaluate these recommendations, students with little to no scientific training were not able to identify the flaws in the evidence compared to students who were later in their college careers and had social science backgrounds. Burrage's (2008) comparative study also found that when asked to discuss a set of research studies that contained methodological flaws, seniors provided more evaluations of evidence and were better at recognizing when studies contained methodological flaws when compared to freshmen.

A recent large-scale study by Arum and Roska (2011) provides one of the most comprehensive studies of students' critical thinking development using open-ended questions in which students' responses are scored based on the quality of their reasoning. In their study, Arum and Roska used the Collegiate Learning Assessment (CLA), which assesses broad abilities important for critical thinking, like the ability to articulate complex problems, justify ideas, and evaluate claims and evidence. The CLA consists of three critical thinking tasks: a performance, a make-an-argument, and a critique-an-argument task.

The performance task attempts to simulate real-life scenarios that require students to evaluate and draw conclusions from several sources of information. One of these scenarios, for example, presents students with the following dilemma: a company was in the midst of purchasing a small private plane, but the model they were interested in purchasing recently crashed. Students are asked to examine different sources of information, some which was relevant and some which was not, and provide a written memo addressing whether the plane had some safety issues, whether there were other possible reasons the accident occurred, and whether they would recommend purchasing the plane.

In the make-an-argument task, students are asked to state their perspective about a number of issues, like whether more government funds should be spent on preventing crime than enforcing crime, and provide relevant reasons to support their position. For the critique-an-argument task, students are asked to read arguments about everyday topics and evaluate the soundness and reasonableness of the argument's logic. As an example, one of the arguments reads,

> The number of marriages that end in separation or divorce is growing steadily. A disproportional number of them are from June weddings. Because June wedding are so culturally desirable, they are often preceded by long engagements as the couple wait until the summer months. The number of divorces increases with each passing year, and the latest statistics indicate that more than 1 out of 3 marriages will end in divorce. With the deck stacked against "forever more" it is best to take every step possible from joining the pool of divorcees. Therefore, it is sage advice to young couples to shorten their engagements and choose a month other than June for a wedding (CLA, 2006).

These tasks are scored using a holistic method, in which a trained researcher uses a Likert-type scale to evaluate multiple dimensions of a students' response, like whether they drew appropriate conclusions, used relevant information, and recognized strengths and weaknesses in information. These scores range from 1-6 where 1 reflects responses that are under-developed or uninformative and 6 reflects responses that clearly identify important facts and ideas that support or refute an argument or justify a claim. However, given that each prompt emphasizes a specific critical thinking skill, responses are also scored according to these specific skills, like whether a student identified a crucial piece of information when critiquing and argument. An overall score is obtained by averaging the scores on the performance and one of the two argument tasks (students' completed either the make-an-argument or critique-an-argument task).

Using this measure of critical thinking, Arum and Roska found that students' CLA scores improved by only 0.18 standard during the first two years in college. Additionally, 45% of the students showed no change in their CLA scores across these two years. Although these findings raise concerns regarding whether college students are learning to think critically in college, some caution is needed when interpreting these findings.

Since Arum and Roskas' goals were to provide a large and representative account of college students' critical thinking development, it was useful for pragmatic purposes to define critical thinking as an overarching set of skills, such as the ability to articulate problems, justify ideas, and evaluate claims and evidence. The CLA was also adequately suited to capture how these skills were utilized in everyday contexts, since these tasks were situated in some practical problem. Given the large scale of this study, using a single score to represent students' depth of evaluation was efficient for data analysis purposes and allowed the authors to draw general conclusions. This method of scoring, however, may have hid some important developmental differences in students' ability to think critically between the different tasks in the CLA (articulating problems, justify ideas, and evaluate claims and evidence). It may also be premature to assume from the findings that students developed little critical thinking skills, since the study only followed students during their first two years in college. Despite these general limitations, the CLA fills an important gap in the critical thinking research because it approached critical thinking in terms of the extent to which students could provide an in-depth response.

Thinking dispositions and their relation to critical thinking development

Being able to reason scientifically about claims is an important critical thinking skill, since it provides individuals with the necessary skills and strategies to examine evidence in relation to the claim. However, scholars have also argued that is it simply isn't enough for individuals to have domain-general scientific knowledge and that equal emphasis should be given to understanding more dispositional aspects of critical thinking (Bensley, 2010; Halpern, 1998; Norris, 1989; Stanovich, 2009). Thinking dispositions refer to two characteristics: a) an individual's willingness to engage in cognitive tasks, and b) the general tendencies an individual exhibits when thinking and reasoning (Halpern, 1999; Stanovich & West, 1998). Much of the psychological research on thinking dispositions has focused on the Need for Cognition and Actively Open-minded Thinking. Below I discuss these constructs and highlight studies that have examined their relationship with college students' critical thinking development.

Need for Cognition. Cacioppo and Petty (1982) proposed that people vary in their *Need for Cognition*, their disposition to engage in and enjoy effortful thinking. Whereas some individuals prefer to think as little as possible in situations requiring effortful processing, others are more motivated to engage in the thinking process. Cacioppo and Petty's (1982) study examined how differences in college students' need for cognition related to whether they enjoyed complex thinking by placing them in either a simple task or complex task condition. For this task, participants were presented with a notebook containing 3,500 random numbers and were instructed to circle numbers based on either an easy rule set (e.g., circle all 1, 5, and 7s) or a complex one (e.g., circle all the 3s, any 6 that preceded 7, and every other 4). They were also given the Need for Cognition scale, a 45-item survey which assessed their willingness to engage (e.g., I find satisfaction in deliberating hard and for long hours) or avoid thinking (e.g., Thinking is not my idea of fun). Participants with a high need for cognition reported that they enjoyed the complex task over the simple task when compared to those with a low need for cognition.

Actively Open-minded Thinking. The work by Stanovich and West (1997) on *Actively Open-minded Thinking* has shown that individuals also vary according in their willingness to listen and consider views that are different from their own. For instance, people with more open-minded dispositions never rule out the possibility that their views may be incorrect. Therefore, when presented with evidence that challenges their beliefs, they are more likely to decouple their personal views and beliefs from when evaluating evidence. In contrast, those with closed-minded dispositions operate with rigid sets of beliefs, which in turn makes them less willing to consider different viewpoints. Thus, people with closed-minded dispositions are more likely to engage in ‘belief-bias’ or ‘myside-bias’ reasoning, in which belief-enhancing evidence is evaluated more favorably than belief-threatening evidence. Using the Actively Open-minded Thinking Scale, a 41-item survey which assessed their level of openness to different viewpoints (e.g., “I believe that different ideas of right and wrong that people in other societies have may be valid for them”), Stanovich and West (1998) showed that individuals with more open-minded dispositions were better able to avoid belief-bias reasoning when asked to evaluate belief-threatening evidence. Interestingly, the work by Stanovich and colleagues has also shown that intellectual ability, as indexed by ACT scores, verbal intelligence, and fluid intelligence measures, are unrelated to avoiding belief-bias reasoning (Macpherson & Stanovich, 2007; Stanovich, 1999, 2009; Stanovich & West, 2008). This demonstrates that individual differences in actively open-minded thinking may be more important for understanding the reasoning process, especially when thinking about I'll than traditional measures of intellectual ability.

The relationship between thinking dispositions and college experience. Some work has demonstrated that thinking dispositions change over the course of college, but only moderately and in a non-sequential pattern (Nelson Laird, 2005; Stewart & Dempsey, 2005). Giancarlo and Facione's (2001) longitudinal study, for example, followed 147 college students at the start of their freshmen year then again towards the end of their senior year. They measured students' thinking dispositions by using the California Critical Thinking Disposition Inventory, a Likert-type questionnaire that captures seven characteristics, two which are similar to the need for cognition and actively open-minded thinking scales. These characteristics were, 1) truth-seeking (intellectual honesty in seeking knowledge), 2) open-mindedness, 3) analyticity (alertness towards situations that require critical thinking), 4) systematicity (being organized and diligent), 5) critical thinking self-confidence (trust in one's own critical thinking abilities), 6) inquisitiveness (intellectual curiosity), and 7) maturity of judgment (ability to see complexity in problems).

When aggregated into a single overall score, students' critical thinking dispositions improved over the course of four years. But when examining each critical thinking disposition, only truthseeking and critical thinking self-confidence improved significantly. Giancarlo and Facione also analyzed differences in students' rate of growth (decrease, maintenance, increase) and found that most students maintained or showed modest gains in these thinking dispositions. Only a few students showed drastic increases or decreases. Facione and Giancarlo additionally conducted a cross-sectional study of freshmen and seniors to examine whether students' critical thinking dispositions were different across sex, class standing, and academic major (Giancarlo & Facione, 2001). Female students scored significantly higher than males on the overall disposition score as well as the open-mindedness and the maturity of judgment sub-scores. However, both males and females had equally low response scores for truthseeking, which reflected students' overall hesitation towards setting aside one's beliefs and values in pursuing truth. When comparing these dispositions by years, seniors had higher overall scores, higher truth seeking, and higher critical thinking self-confidence. Finally, academic major was related to the type of critical thinking skills students developed in college. Business and math/science majors had low disposition scores for truth seeking and open-mindedness when compared to majors in the humanities, arts, and behavioral sciences. What is revealing about Facione and Giancarlo's findings is that, in addition to teach students domain-specific and domain-general critical thinking skills, students' major may also play a role in the thinking dispositions students develop while in college.

The relationship between thinking dispositions and reasoning. Using a wide variety of well-structured and ill-structured tasks, studies have demonstrated that a higher need for cognition is positively related to greater reasoning outcomes (Burrage, 2008; Nair & Ramnarayan, 2000; Nussbaum, 2005; Nussbaum & Bendixen, 2003; See, Petty,

& Evans, 2009). However, most of this area of research has been done under the context of bias reasoning. These studies have generally found that higher need for cognition and actively open-minded thinking are associated with avoiding belief-bias reasoning (Kardash & Scholes, 1996; Toplak & Stanovich, 2002; West, Toplak, & Stanovich, 2008).

In Cacioppo, Petty, and Morri's (1983) study, for example, college students read and evaluated a policy statement advocating for increasing college tuition, which was intended to activate students' biases. For half of these students, the statement contained gave a weak argument for increasing tuition (e.g., improving the physical appearance of the school), whereas the half read a statement containing a strong argument (e.g., citing evidence that showed faculty were leaving for better paying jobs). After reading the policy statement, students were asked to state their attitudes towards the recommendation to raise tuition (1 = negative attitude, … 9 = positive attitude) and completed the need for cognition scale. For the strong argument condition, students with a high need for cognition showed more favorable attitudes towards the recommendation compared to students with a low need for cognition, suggesting that differences in students' need to seek out cognitive challenges related to their tendency to avoid bias responding.

The relationship between thinking dispositions and the ability to evaluate research evidence has been less clear-cut, however (Klaczynski & Robinson, 2000; Macpherson & Stanovich, 2007). Macpherson and Stanovich's (2007) study, for instance, found some contradictory results between students' thinking dispositions and their ability to evaluate evidence. In their study, 195 college students were given three critical thinking tasks that assessed their syllogistic reasoning, argument evaluation, and experiment evaluation skills. The syllogistic reasoning task asked students to state whether a set of syllogisms were logically valid or invalid (e.g. All flowers have pedals; roses have pedals; therefore roses are flowers - which is logically invalid). For the argument generation task, students were asked to write down their arguments about two statements: increasing college tuition (belief-threatening issue) and file-sharing over the Internet (belief-supporting issue). In the experiment evaluation task, students were presented with two experiments: one study which found stay-at-home mothers raised more socially responsible children than working mothers (belief-supporting) and another study which found that stay-at-home mothers raised less socially responsible children (belief-threatening).

Both of the belief-supporting and belief-threatening studies contained experimenter bias. Additionally, the belief-supporting study contained an additional confound (the time children were tested was different for the experiment and control group), which further diminished the validity of the findings. Students evaluated each experiment by responding to three questions: 1) How strongly is the conclusion supported by the results of the experiment?, 2) What is your overall evaluation of the quality of this experiment?, and 3) How persuasive is this experiment? Students responded to each question using a six-point Likert-type scale ranging from unfavorable to favorable evaluations.

Half of the students received decoupling instructions, which asked them to put aside their personal beliefs and think about both sides of the issue. The other half received non-directive instructions, in which they were simply asked to take their time to read about the issues. After completing the reasoning tasks, they were given the short form of the Wechsler Abbreviated Scale of Intelligence (WASCI) consisting of the vocal and matrix reasoning subtests, the Need for Cognition scale, the Actively Open-minded Thinking scale, and questions that assessed their prior beliefs about the topics presented in the argument generation and experiment evaluation tasks.

All three critical thinking tasks produced biased reasoning outcomes. Those who believed that stay-at-home mothers raised more socially responsible children gave more favorable experiment evaluation scores to the supporting study compared to the belief-threatening study. For the instruction set condition, participants who received decoupling instructions displayed less bias in the syllogistic reasoning and argument generation task, but this did not occur for the experiment evaluation task. In addition, the WASCI, Need for Cognition, and Actively Open-minded Thinking scores were positively correlated with correctly answering the inconsistent syllogisms. As they expected, the WASCI was not correlated with avoiding myside bias in the argument generation and experiment evaluation tasks. But different from their expectations, they only found one relationship between the thinking dispositions on the argument generation and experiment evaluation task. This relationship contradicted their prediction, however; a higher need for cognition was related to providing more myside bias responses in the experiment evaluation task.

So why didn't need for cognition and level of open-mindedness help students avoid biased reasoning in this context? It is possible that scientific knowledge is an important requisite for being able to think objectively about experimental evidence. That is, if students do not have adequate domain-general scientific knowledge, then those with a high need for cognition may spend more mental energy basing their evaluations on how well the experimental findings fit with their beliefs and experiences. Macpherson and Stanovich did not provide information regarding participants' majors, class standing, or level of scientific knowledge, so it is unknown how these factors related to students' ability to avoid bias reasoning when evaluate evidence.

The role of prompts, instructions, and college experience on reducing myside bias. Other studies, however, have demonstrated that teaching students scientific principles of research helps reduce bias reasoning (Leshowitz, DiCerbo, & Okun, 2002; Leshowitz & Okun, 2011). In Leshowitz & Okuns' (2011) study, 149 college students with little scientific training were asked to read a court case involving a college coach who was blamed for the death of a player. This case described the coach and the state of college sports in an unfavorable light, which was intended to influence students' favoritism towards the plaintiffs. The plaintiffs (the player's parents) argued that the coach knew the son was taking steroids but did nothing to intervene. In arguing their case, the plaintiffs presented a medical expert who testified that the steroids caused their son to commit suicide. His conclusions, however, were based on his own personal feelings and not on any direct evidence. The medical expert also bolstered his testimony by presenting a set of research studies linking steroids to suicide, which were experimentally flawed and unrelated to the case.

After reading the case transcript, students were asked to offer a verdict (negligent vs. non-negligent). Before reading the case, however, half of the students received a brief lesson on scientific methodology, which discussed the role of control groups, the idea that correlation does not imply causation, and the role of direct evidence in supporting assertions. They found that the brief lesson increased students' skepticism towards the evidence, which in turn helped them reach more non-negligent verdicts (51%) than the control group (35%). It is still important to point out, however, that half of the students in the instructional condition gave a negligent verdict. Although instructions improved scientific reasoning, a high proportion of students in this condition had difficulty applying this reasoning strategy in this context.

Another line of work has also found that simply asking students to reason using a logical person's perspective reduces biased reasoning (Amsel et al., 2008; Klaczynski, 2001b; Thompson, Evans, & Handley, 2005). Although these instructional interventions decrease biases, these results also show that improving how students' reasoning is not an immediate process. For instance, Mil, Grey, and Mandel (1994) found that students who enrolled in an applied statistics course gained few scientific reasoning skills by semester's end when compared to students in a humanities course. In other studies, students were still prone to provide biased responses even after being instructed about biased reasoning (Follmer, Semb, Colombo, & Schreiber, 2004; Pollard, Newstead, Evans, & Allen, 1994). Thus, short interventions, whether in the form of instruction or prompts for critical thinking, have shown to be effective in helping students avoid belief-bias reasoning, but the actual process of developing students' decoupling skills may be a longer and more drawn-out process.

Toplak and Stanvoich (2003) have shown that years in college is associated with using less bias reasoning. In their cross-sectional study, college students were given an argument generation task in which they were presented with three controversial issues: increasing tuition costs, permitting the sale of human organs, and doubling gas prices to discourage driving. They were first asked to rate their position using a 6-point Likert-type scale, then after completing several unrelated tasks, they were asked to generate arguments for and against their own position for each issue. The number of arguments that endorsed their beliefs (myside arguments) compared to the number of arguments that did not (other-side arguments) as used as an index of myside bias reasoning. For all three issues, participants gave significantly more myside than other-side arguments, but myside bias decreased across years in college. After controlling for age and cognitive ability, regression analysis revealed that the number of years in college was an independent predictor of avoiding myside bias.

In sum, higher levels of need for cognition and open-minded thinking are clearly related to avoiding biased reasoning for a variety of tasks, but whether this can be said for tasks requiring evidence evaluation skills is less certain. What is more certain, however, is that domain general scientific knowledge , through the form of short instructional interventions, helps individuals evaluate evidence-based claims and arguments. Toplak and Stanovichs' (2003) work has also helped demonstrate that years in college is related to avoiding myside bias reasoning. However, further work is needed to establish how college exposure is related to developing students' thinking dispositions, scientific knowledge, and their ability to provide unbiased scientific evaluations.

Dual process models of reasoning

Dual process models of reasoning characterize mental functioning as stemming from two types of cognitive processes: the autonomous (Type 1) and the non-autonomous (Type 2) (Evans, 2003; Evans & Frankish, 2009; Stanovich, 1999). Type 1 reasoning is fast, automatic, intuitive, and unconscious, and is primarily constructed via associative learning. This type of reasoning does not require much cognitive effort and is seldom guided by extensive analysis. Type 2 reasoning, on the other hand, is deliberate, effortful, and controlled.

Although dual process models share similarities with the work on thinking dispositions, rather than focus on an individual's motivation towards thinking or level of open-mindedness, this perspective assumes that an individual's thinking functions within the scope of basic autonomous and non-autonomous cognitive systems. In the context of evaluating information and evidence, Type 1 reasoning is characterized as an intuitive-experiential process, in which individuals naturally defer to their personal beliefs, experiences, heuristics, or intuitive feelings. Type 2 reasoning is characterized as a rational-analytic process, in which individuals consciously process information by using some principle of logic or reasoning.

Research on dual process models primarily emerged from Kahneman and Tversky's work who showed that individuals often relied on heuristics when drawing inferences from information (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). Their pioneering work found that participants often ignored base-rate information (70 engineers, 30 lawyers) when deciding whether an individual from the sample was an engineer or lawyer. Instead, participants based their decision on a quick heuristic shortcuts; if the individual's personality resembled a lawyer more that it did an engineer, then the individual was probably a lawyer (Kahneman & Tversky, 1973).

One assumption about dual process models is that they function in an interactive parallel (De Neys, Vartanian, & Goel, 2008; Evans, 2007). That is, even when thinking intuitively, individuals are still capable of noticing when they are being biased (Epstein, 1994). Another assumption of dual process models is that both systems of reasoning are optimized for different situations (Stanovich, 2010). For instance, intuitive-experiential processing works optimally in situations where time and knowledge are limited (Gigerenzer & Goldstein, 1996; Gigerenzer, Hertwig, & Pachur, 2011; Goldstein & Gigerenzer, 2002). Kahneman & Klein (2009) have also noted that years of effortful training in a given domain can lead to developing 'intuitive expertise' where an expert's quick and intuitive judgments are just as optimal, if not better, than the effortful judgments of non-experts. Scholars assert that since experts have obtained a deep knowledge base and experience in a given domain, they become skilled at noticing different problem structures and applying effective heuristic strategies to solve them (Chi, Glaser, & Reese, 1982; Larkin, McDermott, Simon, & Simon, 1980a). In one line of research, physics students had difficulty solving physics problems despite having the necessary background knowledge to solve such problems. In contrast, physics experts were able to quickly and accurately identify the theoretical principles underlying each problem and solved these problems with greater ease (Larkin, McDermott, Simon, & Simon, 1980b). Studies examining chess players have also found that, under a variety of situations and task demands, chess experts are faster at recognizing important chess board patterns and better at choosing optimal moves when compared to novice chess players (F. Gobet & H. Simon, 1996; F. Gobet & H. A. Simon, 1996). In one notable study, the quality of chess experts' moves remained high even after the experimenters drastically reduced the time they could spend on each move, supporting the idea that developing expertise can lead to obtaining effective heuristic-based strategies (F. Gobet & H. A. Simon, 1996).

However, other studies have found that intuitive reasoning can lead to making biased judgments, especially in situations where individuals evaluate evidence that is consistent or inconsistent with their prior beliefs and expectations (Leshowitz & Okun, 2011; Macpherson & Stanovich, 2007). Similar to the work on thinking dispositions, the work on dual process models of reasoning have also focused on examining biased reasoning processes. However, this area of research has not fully considered how this in the context of evaluating evidence. Instead, much of this work has focused on using deductive reasoning (Amsel et al., 2008; Evans, 1998) or syllogistic reasoning paradigms (e.g., No addictive things are inexpensive; some cigarettes are expensive; therefore, some addictive things are not cigarettes – which is an invalid conclusion) to observe differences in individuals' intuitive-experiential versus rational-analytic processes (De Neys, 2006; Evans, Handley, & Harper, 2001). One study by Klaczynski, Gordon, and Fauth (1997) does provide some evidence, showing that the tendency to use either intuitive-experiential or rational-analytic processes is to patterns in college students' evaluation of evidence. In their study, college students were given a series of summaries to read that were relevant to their academic major. These summaries described hypothetical situations in which people made arguments that were based on an observation of small samples, on questionable correlational relationships, an on a poorly designed experimental studies. The arguments, however, were either goal-enhancing (accountants are smarter), goal-neutral, or goal-threatening (accountants have poor marriages).

They were asked to rate the quality of the arguments in the study on a 9-point scale. Additionally, students provided a written explanation describing why they believed the arguments were valid or invalid. These written explanations were assessed for depth of scientific evaluations (0 = no mention of statistical concepts, 1 = poorly elaborated statistical concepts, 2 = detected law of large numbers, covariation comparison, experimental flaw). Participants also completed the Rational-Experiential Inventory (Epstein, Pacini, Denes-Raj, & Heier, 1996), which assessed their individual preference using more intuitive-experientially oriented reasoning versus rational-analytic reasoning. This inventory was composed two sub-scales: 1) Need for Cognition and, 2) Faith in Intuition scale. Whereas the Need for Cognition scale measured a individual's willingness to engage and enjoy cognitive activities, the Faith in Intuition scale captured an individual's reliance on using intuition (e.g., I hardly ever go wrong when I listen to my deepest feelings to find an answer) or some form of logic (e.g., Using logic usually works well for me in figuring out problems in my life) when engaged in thinking.

In all, students thought the goal-enhancing arguments were more persuasive and valid than the goal-threatening arguments. Students' depth of scientific evaluations were also lower when the arguments were goal-enhancing or neutral. Students' rational-experiential dispositions were also related to this process; those who reported a greater preference for thinking rationally provide more in-depth and less biased evaluations than those who preferred to think more experientially.

As demonstrated by Klaczynski et al. (1997), dual process models of reasoning provide a useful framework for understanding how students evaluate claims and arguments. In line with previous work using different reasoning tasks, they found that students engaged in belief-bias reasoning. Their study added to our understanding, however, by directly observing the relationship between students' belief-bias reasoning and the depth of their evaluations. When arguments were goal enhancing, students did not spend much effort evaluating the evidence. But when students encountered goal-threatening arguments, they critiqued the evidence by mentioning statistical concepts and detecting the experimental flaws. The most revealing aspect of their work found that students who reported a greater preference for think rationally were better able to avoid belief-bias reasoning when evaluating goal-threatening arguments.

Dual process models of reasoning may help scholars understand whether college influences how students utilize these two cognitive systems when thinking critically. Given that scientific knowledge is an important requisite for being able to evaluate evidence, it would be interesting to understand whether differences in students' scientific knowledge are associated with a greater preference for thinking more rationally. Additionally, since other dispositional factors, like open-mindedness, are associated with avoiding biased reasoning, it would also be essential to examine whether a preference for thinking more rationally is also associated with greater open-mindedness. Finally, developmental studies would help us understand whether college training, especially courses that emphasize scientific research methods, helps students shift away from using less intuitive-experiential and more rational-analytic processes (like scientific reasoning) when evaluating evidence.

One issue regarding dual process models is that studies have yet to directly observe students' intuitive-experiential responses. The work on dual processing have primarily measured biased reasoning by counting the number of instances individuals provide either a biased or unbiased responses. It is also common to use students' ratings of evidence to reflect whether they are thinking more or less biasedly. Although these methods are economically feasible, over-relying on them may obscure how we understand the reasoning processes. In order to shed light on how students use either cognitive system for reasoning, work is needed that observes what students say when describing or justifying their responses. Indeed, the work by Klaczynski and colleagues (1997) has done this, and showed that more rationally oriented individuals were using more scientific reasoning when evaluating evidence. However, measures that assess students' depth of scientific evaluations do not provide information regarding what students are doing when *not* reasoning scientifically. One promising way to understand how students use both reasoning systems is to observe when either instance occurs. For instance, Sá et al. (2005) found that when participants were asked to generate hypothetical evidence regarding several issues (what causes prisoners to return to crime after they are released?) they tended to discuss their personal experiences and opinions more so than they discussed the importance of obtaining some form of evidence. From a developmental perspective, this method of observing reasoning allows us to make clear descriptive comparisons regarding whether college students' experiential versus scientific responses change over the course of college.

The relationship between epistemic beliefs and critical thinking

Although dual process models provide a well-informed and well-supported basis for understanding human reasoning, other psychologists have focused on understanding how epistemic beliefs play a role in shaping peoples' reasoning processes (Hofer & Pintrich, 1997; King & Kitchener, 1994; Sandoval, 2005). Epistemic beliefs refer to how people come to understand what they know about the world, in terms of where knowledge comes from and how knowledge is justified. As Hofer and Pintrich note, one line of work has focused on capturing age-related changes in the assumptions people hold about reality and the certainty of knowledge. The other line of work has tried to uncover how epistemic beliefs shapes peoples' reasoning processes and their ability to think critically (Hofer & Pintrich, 1997). In terms of how epistemic beliefs are related to critical thinking, as Stanovich (2009) and others point out, it is one thing for students to use critical thinking in the service of instrumental rationality, where one thinks critically in order to fulfill some personal goal. It is another thing to use critical thinking in the service of epistemic rationality, where one tries to ensure that their beliefs and values represent an accurate reflection of what is known about the world (Baron, 2008; Foley, 1987; Stanovich, 2009; Stanovich & Stanovich, 2010; Sternberg, 2002).

The work on epistemic development was primarily born out of William Perry's (1970) interviews with male college students regarding their educational experiences. As freshmen, the students in Perry's study perceived their experiences in dualistic positions; things were either right or wrong. But when interviewed again as seniors, students focused more on considering the role context and understanding different points of view. Belenky et al.'s (1986) study of female college students also found changes in their epistemic development. Initially, students viewed their educational experiences as silent learners, where knowledge was not obtained by participating but by simply listening to authority figures. But as they became older, they began to view the learning process something that was personally constructed by an individual. Both Perry and Belenky et al.'s studies were influential in showing that students' thinking reflected a developmental process, where the assumptions they used to interpret their experiences became increasingly complex as they grew older.

Rooted in Perry's ideas, King and Kitchener (1994; 1981) developed a theory of post-adolescent epistemic development, known as the reflective judgment model. This model describes how individuals understand reality and knowledge as they grow older. This is observed by examining what assumptions people have about knowledge when asked to justify their views. Kitchener and King developed the Reflective Judgment Interview to measure the assumptions people hold about the certainty of knowledge, how knowledge is acquired, and how knowledge is justified. This interview presents people with four dilemmas that describe contradictory arguments. The following example represents a dilemma about the safety of food additives.

> There have been frequent reports about the relationship between chemicals that are added to foods and the safety of these foods. Such studies indicate that such chemicals can cause cancer, making these foods unsafe to eat. Other studies, however, show that chemical additives are not harmful, and actually make foods containing them more safe to eat (p. 260, King & Kitchener, 1994).

After reading the following passage, the interviewer prompts the participant to endorse a point of view, which is then followed up by six follow-up questions asking participants to justify how they came to adopt this view and whether they can ever know if their position is correct. Kitchener and King propose that the way people justify their views represents seven stages of epistemic reasoning. Stages 1 through 3 represents *pre-reflective reasoning*, where knowledge is viewed as being either correct or incorrect and obtained by some direct experience or through some higher authority.

People who reason pre-reflectively tend to justify their views by appealing to what feels right at the moment, by citing some direct example or observation, or by referring to a higher authority. People who reason this way do not identify evidence or the accumulation of evidence as important for understanding reality. Instead, they portray knowledge as something that is already known, which can color how they view complex problems. As King, Wood, & Mines note, "people who hold these assumptions cannot differentiate between well- and ill-structured problem, all problems as though they were defined with a high degree of certainty and completeness" (King, Wood, & Mines, 1990, p. 169).

Peoples' reasoning in stages 4 through 5 reflects an appreciation for evidence in justifying knowledge claims. Although this represents a major shift in thinking, people struggle understanding whether knowledge claims can ever be known due to what they perceive as major limitations in obtaining evidence, which reflects a form of *quasi-reflective reasoning*. Under this view, knowledge cannot be justified until all the necessary evidence has been obtained, but since it is practically impossible to do so, people can never know what is true. Additionally, since people are inherently biased in how they accumulate and interpret evidence, knowledge is seen as being highly subjective. In general, people who reason quasi-reflectively are conservative about judging whether one subjective viewpoint is better than another. Instead, multiple realities can exists, so what is true for one individual may not be true for another. Although quasi-reflective approach still reflects some uncertainty about whether knowledge can ever been known, a basic appreciation of evidence and the need to examine claims provides the necessary framework for being able to evaluate and appreciate opposing realities.

Stages 6 through 7 represents *reflective reasoning*, where people acknowledge the inherent difficulties involved in evaluating knowledge claims, but understand that some claims are more valid than others. Reality can be determined by comparing the evidence that each perspective presents, in terms of whether the evidence is relevant to the claim that is being made, whether the evidence was obtained through reasonable means or well-regarded principles of inquiry, and whether the evidence is being interpreted correctly. People who reason reflectively can also understand the changing nature of knowledge, in that knowledge can be revised as more evidence becomes available.

Kitchener and Kings' (1994) 10-year longitudinal study found that, for over ninety percent of their sample, individuals' epistemic assumptions followed a steady developmental progression, where adolescents generally endorsed absolutist views of reality and knowledge claims but began understanding the nature of knowledge as being more complex as they grew older. In one cross-sectional study, Kitchener and King (1981) interviewed high school juniors, college seniors majoring in liberal arts fields, and doctoral students studying liberal arts fields. Ninety percent of high school students' reflective judgment scores were between stages 2 and 3, suggesting that they justified their beliefs using absolutist assumptions about knowledge. Eighty-five percent of college students' scores were between stages 3 and 4.5, which reflected their growing appreciation for evidence as well as their apprehension to view reality and knowledge as certain. Graduate students showed the highest level of epistemic understanding, where seventy-five percent scored above stage 5. In contrast to high school and college students, graduate students understood reality and knowledge claims as something that could be rationally justified through evidence. Kitchener and King were also able to show that other competing factors, like sex, and verbal ability could not account for the difference in epistemic scores, suggesting peoples' epistemic understanding represents a unique developmental process.

King and colleagues also examined whether the development reflective judgment varied by academic training. Since the Reflective Judgment Interview involves justifying one's views about ill-structured problems, King, Wood, and Mines (1990) hypothesized that students in social sciences related fields would be better trained to think about ill-structured problems than students from other fields. To test this, they interviewed forty college seniors and forty graduate students who studied either social science (sociology and psychology) or math. In addition to the Reflective Judgment Interview, students were also given the Watson Glaser-Critical CTA and the Cornell Test of Critical Thinking, which measured their ability to solve well- and ill-structured logical problems.

For the Reflective Judgment Interview, they found significant differences by education level, where college seniors had lower reflective judgment scores compared to graduate students. Although college seniors' reflective judgment scores did not differ by training, graduate students in social science had higher scores than graduate students' in math. For the Watson Glaser and Cornell tests, college seniors had lower scores than graduate students. College seniors in math, however, performed better than seniors in social science. But interestingly, both graduate students in math and social science performed similarly. These findings suggested that different undergraduate discipline promote specialized critical thinking skills, but did not have a special influence on students' reflective judgment.

In order to account for the role of education on epistemic development Kuhn's (1992) comparative study, which used a paradigm similar to the Reflective Judgment Interview, sampled a group of adolescents, young adults, and older adults. For the young and older adults, half were college educated, while the other half were not. Participants were asked to justify their theories regarding everyday problems, like what causes criminals to return to crime. When probed by the interviewers to support their argument (*How do you know that this is the cause? What evidence can you give to show this?*) participants who attended college were more likely to cite the importance of evidence when justifying their theories when compared to participants who had not attended college. Of the young adult sample, 80% of the college-educated participants cited evidence as important where only 35% of non-college participants did so.

Another approach to studying epistemic development have focused on the learning process, in terms of how students understand how knowledge should be acquired as learners (Schommer, 1990, 1993; Schraw, Benedixon, & Dunkle, 2002). For instance, Baxter Magolda's (1992) longitudinal study primarily examined how college students understood the role of the individual, instructors, and peers in the learning process during their time in college. They were able to show that students' assumptions about learning also followed a developmental progression, where they initially viewed learning as absolute, in that the knowledge they acquired was certain, but then moved to understanding learning as a contextual process as they matured, where they grew more considerate of understanding competing viewpoints and context.

Using this 'beliefs about knowledge' / 'beliefs about learning' distinction, Schommer-Aikins (1990) has argued that people have a number of different kinds of epistemological beliefs, which are more or less independent from each other. In an effort to demonstrate this construct, Schommer-Aikins developed the Epistemic Beliefs Questionnaire, a 12-item Likert-type questionnaire, which assessed four types of epistemic beliefs: 1) Simple Knowledge (knowledge as isolated facts), 2) Certain Knowledge (knowledge as absolute), 3) Innate Ability (learning as innate), and 4) Quick Learning (learning as quick or no-at-all). Schommer-Aikin's Epistemic Beliefs Questionnaire provided one of the earliest quantitative assessments of epistemic development while also emphasizing the importance of assessing *beliefs* in this area of research.

This learning-based domain of epistemic development, however, has been criticized for being conceptually flawed (Hofer & Pintrich, 1997, 2002; Sandoval, 2005, 2009). This is primarily because beliefs about learning are self-theories, theories about how an individual views their own personal attributes (Dweck, 2000; Dweck, Chiu, & Hong, 1995). For example, in Schommer-Aikin's Epistemic Beliefs Questionnaire, the statements assessing Innate Ability actually capture individuals' implicit theories about their intelligence, whether they think their own intelligence is innate (e.g., *The really smart students don't have to work hard*) or malleable (e.g., *Genius is 10% ability and 90% hard work*). This is also the case for the items assessing Quick Learning, which assesses an individual's implicit theories about their own learning abilities, whether they think it is fixed (e.g., *Going over and over a difficult text book chapter usually won't help you understand it*) or incremental (e.g., *If I find the time to reread a textbook chapter, I get a lot more out of it the second time*). In order to avoid confusion surrounding constructs dealing with self-theories, Hofer and Pintrich recommend that the study of epistemic beliefs be conceptually defined within the boundaries of how individuals understand both *the nature of knowledge* and the *process of knowing* (Hofer & Pintrich, 2002). The nature of knowledge refers to how an individual understands what counts as knowledge. This is also related to how individuals understand the certainty and simplicity of knowledge. The process of knowledge, on the other hand, refers to the how an individual understands where knowledge comes from and how knowledge is justified.

More recent work in this field has begun to appreciate the idea that people may have distinct epistemic beliefs about different domains of knowledge and that these beliefs can differ cross-culturally (Karabenick & Moosa, 2007; Muis, Benedixon, & Haerle, 2006; Pintrich, 2002; Schommer-Aikins, 2004). Estes et al's. (2003) study, for example, examined American and British college students' beliefs about psychological and biological knowledge. Using a Likert-type scale ranging from 1 to 7, students were asked to state whether they agreed or disagreed with 4 statements regarding the certainty of scientific research (e.g., *On most issues in this field, with enough careful research, scientific experts can sooner or later be certain that their findings are correct*). For the psychological domain, these statements focused on the certainty of scientific research on children's social and emotional development, whereas statements for the biological domain focused on children's physical and biological development. Estes and colleagues found that students were less certain about psychological research compared to biological research. When asked to explain why, students generally stated that there were more inherent difficulties in conducting psychological research, in that social and emotional constructs were more complex, encompassed greater individual variability, and were further affected by other variables. American students' however, were more skeptical about both fields that British students were, which may reflect cultural differences in how students understand and value scientific research.

Another line of work has focused on understanding the relationship between individuals' epistemic beliefs and their cognitive processes. (Hofer, 2004; Kardash & Scholes, 1996; Schraw et al., 1995; Weinstock, 2009). One area of this work examines the interrelationship between individuals' beliefs about knowledge and their metacognitive processes, what Hofer (2004) as referred to as "epistemic metacognition." In explaining this relationship Hofer states, *"For example, if one believes knowledge is finite, then multiple sources of information might prove redundant and confusing; a search for competing truths is unnecessary, as is any attempt to resolve those that emerge"* (Hofer, 2004, p. 47). In Hofer's (2004) descriptive study of preliminary data, a group of high school and college students were asked to find information online about "bees and their communication behavior" for a report. Students' epistemic beliefs were assessed by asking them to select their three best sources and explain their choices for doing so. Although Hofer did not report any developmental patterns, she found that the sources students chose for their report reflected their level of epistemic understanding about the certainty and simplicity of knowledge. In one example, a student selected a book from 1908 because "in biology when they know it, it's not likely to change" (p. 53).

In one study, Kardash and Scholes (1996) examined whether college students' epistemic beliefs and need for cognition related to how they interpreted controversial issues. Students were first asked to state whether they believed HIV caused AIDS. Students then read an article summarizing a debate between two research teams, one that supported the view that HIV was the sole cause of AIDS, and the other that HIV did not cause AIDS. The article was intended to be inconclusive in nature; both research teams presented equally strong evidence and rebuttals but the article itself did not conclude who was ultimately correct. Students were asked to write what they concluded from the articles. From there, they completed the Epistemic Beliefs Questionnaire (only the Certain Knowledge component was used for the analysis) and the Need for Cognition Scale. Kardash and Scholes found that the more students viewed knowledge as certain, the more likely they were to write biased conclusions that favored their initial views. The same pattern was found for students' need for cognition, in that students who disliked and avoided the process of thinking were more likely to provide biased conclusions. Conversely, students who viewed knowledge as less certain and who had a higher need for cognition were more likely to draw inconclusive results from the article. This study was important for showing that epistemic views about knowledge and their motivation towards thinking were related to how they justified their views.

How anecdotal stories influence reasoning

At the most basic level, evidence-based claims contain two important components, the assertion and the supporting evidence. They way evidence-based claims are reported in everyday settings, however, seldom provide such clear-cut distinctions. For instance, scientific news articles convey the results of scientific research studies in a clear an accessible ways to their readers. One way they do this, however, is by beginning the article with a compelling anecdotal narrative. Take the following anecdotal narrative for example,

> Amanda knew Frank was shy when she married him. Although Frank is a warm and caring person, his introverted tendencies have started to take a toll on their relationship. At social gatherings, Frank always tries to find a quiet corner. This makes things awkward for Amanda, who doesn't know whether to sit next to him or continue to mingle with others. Amanda remarks, "Although I love him, it frustrates me when he gets like this. It makes me question our compatibility." The latest research is also revealing that introverted partners negatively impact marriages.

Such narratives help readers create a vivid mental images of people in relevant scenarios, which in turn increases readers' ability to comprehend novel information (Bower & Morrow, 1990; Graesser, Singer, & Trabasco, 1994). However, narratives also have an influential effect on individuals' perceptions and judgments. When listening to stories, individuals actively construct and infer causal relationships based the events that transpired. As suggested by scholars, anecdotal narratives can activate experiential systems, where individuals base their judgments on whether a story seems reasonable or believable and not how well a claim is supported by evidence (Epstein, 1994; Nisbett & Ross, 1980; Winterbottom et al., 2008).

In Strange & Leung's (1999) study, for example, people made judgments about the causes of poverty based on whether an anecdotal narratives attributed the problem to the individual or larger structural issues. In a study by Dahlstrom (2008), undergraduate students read a narrative text that contained claims about the natural world (e.g., Jellyfish avoid the fast moving currents of the shallows). These claims were paired alongside either, a) anecdotal stories with causal narrative structures, b) anecdotal stories with non-causal narrative structures, and, c) non-anecdotal text. Dahlstrom found that claims containing causal anecdotal narratives were rated as more truthful than claims with non-casual narrative structures and non-anecdotal texts. Interestingly, other studies find that when anecdotal narratives and statistical evidence are presented alone, anecdotal narratives are rated as less persuasive and less believable than statistical evidence (Baesler & Burgonn, 1994; Hoeken, 2001).

Much of the work on anecdotal narratives comes from medical decision-making studies since anecdotal narratives are commonly used to help patients understand their illness and make decisions about various treatment options. Some work has found that anecdotal information can aid decision making if the information that is presented is balanced, where the advantages and disadvantages of various treatment options are discussed (Bekker, Hewison, & Thorton, 2003). However, in other cases, anecdotal information can reduce individuals' ability to attend to and critically evaluate relevant information (Beyerstein, 2001).

It is not entirely clear, however, whether anecdotal narratives are more influential than scientific evidence. For instance, Winterbottom et al.'s (2008) review of 17 medical decision making studies found that anecdotal information only had an influential effect on a third of the reviewed studies. Winterbottom et al. note that much of the discrepancies may be due differences in how readers perceive different medical topics and/or whether the narrative information is written in first- or third-person. For instance, anecdotal information was found to be more influential than statistical information when the topic was about osteoporosis and was written in third-person. However, when the topic was about organ donation and the anecdotal information was written using a third-person narrative, statistical information was more influential than anecdotal information. Other work has also shown that anecdotal information has little influence over peoples' decisions if statistical information is presented in using simple graphs (Fagerlin, Wang, & Ubel, 2005).

Although some argue that anecdotes, which contain causal narrative structures, influence individuals' reasoning processes (Anderson, 1983; Dahlstrom, 2010), it is difficult draw general conclusions about the role of anecdotes across studies with qualitatively different content. Despite the uncertainty in the literature, however, it is still important to consider whether anecdotes influence how college students perceive and evaluate evidence-based claims. As previously noted, everyday scientific news reports often use anecdotal stories to discuss scientific research studies, which may influence readers to believe the article to be more valid than what the evidence suggests. This leads to the second point, in that not only is it important for educators to understand whether students are developing strong scientific reasoning skills, or whether they are better able to avoid biased reasoning, but also whether they can resist the persuasive allure of anecdotal stories. The ability persuasive information and examine relevant evidence is an important critical thinking skills that has yet to receive extensive attention.

Dissertation Goals and Contributions to the Field

The goal of my dissertation is to further understand the development of college students' critical thinking skills using a sample of underclassmen and seniors majoring in psychology. I situate students' critical thinking in the context of evaluating everyday evidence-based claims. These types of claims are commonly found in the health and science sections of newspapers and online news-sites. The format for these article usually begin with a title (e.g., "Younger Siblings are More Creative") followed by a summary of the supporting study. These types of reports seldom provide information about how a study was conducted, however. In order to make these articles accessible to a wide audience, the methodological details are often omitted, such as the sample characteristics, the measures, and the scoring procedures, to name a few. There are also instances where scientific news reports misinterpret the results of a study. They may over-state the significance and importance of a study when the results of the study relatively weak. They may also draw erroneous causal claims from correlational studies.

Another way news articles convey scientific research studies in a clear an accessible ways is by introducing the topic with a compelling anecdotal narrative. These narratives help explain the significance of the scientific study by helping readers create vivid mental images of people in relevant scenarios. Although anecdotes help facilitate comprehension, they also can lead an individual to base their judgments and decisions on how well a story supports a claim, and not how well a claim is supported by evidence.

Previous studies have relied on multiple choice responses or Likert-type scale items to capture students' critical thinking process. Although these measures help us understand whether students are thinking more or less critically, they do little to inform us about how students reason. My work adds to the literature by observing students' reasoning strategies. I observe students' reasoning strategies by using a dual process model. As reviewed previously, dual process models of reasoning argue that people have two general systems for reasoning: the intuitive-experiential and the rational-analytic system. Where the former is guided by personal intuitions, beliefs, and experiences and the later is guided by some principle or logic or reasoning. In the context of evaluating evidence-based claims, I consider rational-analytic systems to be guided by scientific principles of reasoning. I argue that attending college helps students shift away from using intuitive-experiential forms of reasoning and more towards scientific ones when evaluating evidence. This is especially the case for students who take courses that challenge them to think critically about the relationship between knowledge and evidence. For psychology majors, a major component of their training involves taking courses in research methods and applied statistics, which in turn helps promote their domain-general scientific knowledge and scientific reasoning strategies. Thus, I attempt to capture college students' critical thinking development by examining whether seniors majoring in psychology provide less intuitive-experiential reasoning and more scientific reasoning strategies when evaluating evidence when compared to freshmen students.

In observing intuitive-experiential reasoning strategies, I distinguish between strategies in which students use either their, a) *Opinions & Explanations* or, b) *Beliefs & Experiences* to evaluate evidence-based claims. Opinion and explanation-based strategies are those in which students state their personal views and opinions about a study or provide a personal explanation. In doing so, however, they do not explicitly refer to some personal belief or experience as being the primary source of these opinions and explanations. Belief and experience-based strategies, on the other hand, occur when students draw from their prior beliefs and previous experiences to evaluate the evidence or discuss why they agreed with a study's claim.

There are numerous ways to evaluate evidence scientifically, from examining how a measure was aggregated to considering whether a research design contain inherent biases. Instead of providing an exhaustive list of the various scientific strategies students utilize, I collapse these strategies into three broad categories. The first category, *Internal*

Threats to Validity captures when students point out possible measurement errors or inherent biases in the design of the study that may invalidate study results. *ANCOVA Reasoning,* on the other hand captures when students consider the role of third variables outside of the study that could interact or explain the study's outcome. Finally, *Methodological & Statistical Reasoning* captures when students assess the soundness of a methodological and statistical procedure as well as the strength and impact of a study's outcome. I acknowledge that internal threats to validity fall within the domain of methodological and statistical reasoning. The reason I observe internal threats separately, however, is because my initial coding of Study 1 revealed that students routinely used this specific strategy to evaluate evidence.

Another contribution I make to the work on students' critical thinking development is that I also examine the depth of students' scientific evaluations. Scholars have noted that is isn't enough to observe students' scientific reasoning strategies. It is equally important to capture how well students are able to explain and justify their scientific evaluations, since it can tell us how the level in which students engage in the scientific reasoning process. For instance, two different students may use and the same scientific reasoning strategy when evaluating evidence but their depth of their evaluations may be quite different. Where one students' ANCOVA based evaluation may simply state, "*While researchers found the negative correlation between the two variables, there are other possible influences over social involvement and how it can affect marriage."* another students' evaluation may try to explain why it is important to consider the role of other variables, "*This seems like it could be possible, but it would be important to look at possible mediating variables. Perhaps introverted people are also more likely to suffer* *from disorders such as depression or anxiety. These would be important things to look at.*" Providing an observational account of students experiential and scientific reasoning strategies gives us a general picture of how students approach the reasoning process, while examining the depth of their scientific evaluations will add another important level of understanding – whether students are indeed learning to provide critical and scientifically justified evaluations.

As prior work has shown, there is a close relationship between domain-general knowledge and reasoning abilities. If an individual has few domain-general scientific reasoning skills their ability to provide in-depth evaluations will be limited. However, another line work suggest that thinking dispositions also matter in understanding why some individuals prefer to reason experientially versus scientifically and why some are better able to avoid belief-bias reasoning than others. For instance, differences in need for cognition and has been shown to relate to how much effort people put forward when reasoning, while differences in open-mindedness is related to individuals' ability to decouple their beliefs and experiences when evaluating information. Although these thinking dispositions have been well studied in a variety of thinking tasks, it is not well understood whether they are related to an individuals' ability to evaluate evidence-based claims. My work identifies specific relationships between individual thinking dispositions and the tendency to prefer using either experiential or scientifically based reasoning strategies.

Epistemic beliefs are another important construct that are related to differences in how individuals reason. The work of King and Kitchener has demonstrated that an individuals' level of epistemic understanding is closely related to their ability to solve ill-structured problems. People who reason pre-reflectively think about problems and justify their views in simple, absolutist ways. In contrast, people who reason reflectively view problems with greater complexity and thus better able to examine available evidence when drawing conclusions. Their longitudinal work suggests that an individuals' epistemic beliefs increases in complexity from post-adolescence to adulthood. Their work has also shown that attending college is related to higher levels of reflective-judgment. Although exposure to college seems to facilitate greater epistemic development, not much is known about how this development relates to students' critical thinking skills. I try to examine the relationship between students' beliefs about the certainty social scientific knowledge and their tendency to utilize experiential versus scientifically reasoned evaluations. Furthermore, I compare freshmen with senior psychology majors in order to observe differences in their epistemic development. Finally, I examine the relationship between students' epistemic beliefs and, a) their thinking dispositions, b) their experiential and scientific reasoning strategies, and c) and the depth of their scientific evaluations.

Research Foundations and Hypotheses for Studies 1 and 2

The first study I present provides an observational account of college students' experiential and scientific reasoning strategies. I examined how college underclassman used either strategies when evaluating evidence in two different contexts; one in which they were simply asked what they thought about the studies and another in which they were asked to critically evaluate the studies. In doing this, I wanted to examine whether prompting students to think critically drastically shifted their reasoning processes or whether students still relied on using experience-based reasoning in this context. I also examined the relationship between three individual difference measures (need for cognition, actively open-minded thinking, prior beliefs) on students' experiential and scientific reasoning strategies, and the depth of their scientific evaluations.

For Study 2, I examined students' critical thinking development by comparing college freshmen with seniors majoring in psychology. Like the first study, I observed student's experiential and scientific reasoning strategies as well as their depth of scientific evaluations. I additionally compared whether there were differences between students' thinking dispositions (need for cognition, actively open-minded thinking), epistemic beliefs, and level of scientific knowledge. Students were asked to evaluate a set of evidence-based claims that mimicked scientific news reports. I was additionally interested in understanding whether anecdotal stories influenced students to evaluate the studies more favorably. Therefore, for half of the students in this study read and evaluated claims containing alluring anecdotal stories.

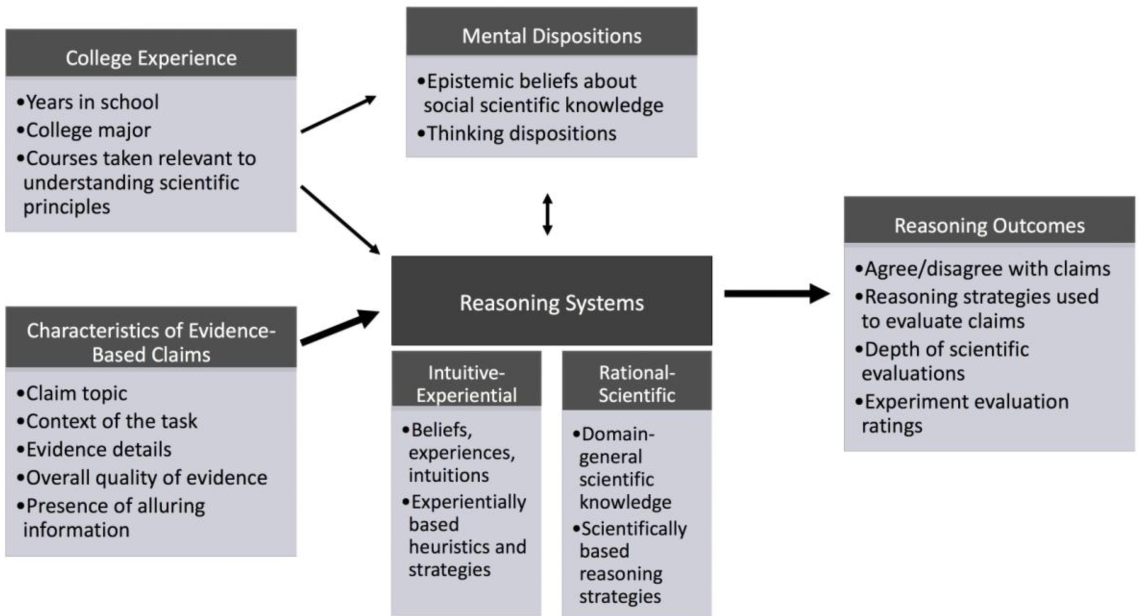
The model I present below (Figure 2.1) illustrates how I conceptualize college students' critical thinking development. Based my current understanding of the literature, I use this model to show how the characteristics of a claim affect students' reasoning processes, while also illustrating how individual characteristics (years in school, major, level of scientific knowledge) is related to how they approach reasoning when evaluating evidence. Finally, I attempt to capture how differences in students' thinking dispositions and epistemic beliefs moderate the reasoning process, in that students who have more rationally-oriented and open-minded dispositions, and who have more reflective epistemic beliefs will be more inclined to use scientific reasoning strategies compared to those with less rationally-oriented and open-minded dispositions, and who have less reflective epistemic beliefs. Based these general assumptions, I predict the following:

1) Given that individuals have a preference for thinking experientially, college students will engage in more experiential reasoning than scientific reasoning when evaluating evidence-based claims. It is only when explicitly instructed to think critically that student will use more scientific reasoning.
2) With regards to the thinking disposition measures, Need for Cognition and Rational-Experiential reasoning will be positively correlated with Actively Open-minded Thinking. Individuals who are more likely to report a higher need for cognition and a greater preference for thinking rationally will also report being more open-minded. Additionally, individuals' epistemic beliefs about social scientific knowledge will be positively related to these thinking dispositions. The more reflective a students' beliefs about the nature of social scientific knowledge, the more likely they will report a preference for thinking rationally and being more open-minded to different viewpoints.
3) When comparing freshmen and senior's individual characteristics, college seniors will report having greater domain-general scientific knowledge than college freshmen. Additionally, seniors will report a greater preference for thinking rationally, being more open-minded, and having more reflective epistemic beliefs than freshmen.

4) College seniors will evaluate evidence-based claims more scientifically than college freshmen. Seniors' scientific evaluations will also be more in-depth than freshmen students' scientific evaluations.

5) Evidence-based claims that contain alluring anecdotal stories will be evaluated more favorably than claims that do not contain anecdotal stories. Given that they trigger experientially based models of causality, anecdotal stories will also influence individuals to evaluate claims using more experiential reasoning than scientific reasoning. College seniors will be less susceptible to alluring anecdotal stories than freshmen, which will be reflected using less experiential evaluations when compared to freshmen.

Figure 2.1. Proposed model of college students' critical thinking development



CHAPTER III

Study 1

The ability to reason scientifically about research evidence is important for critical thinking, since this reasoning process involves evaluating evidence in relation to claims, theories, and hypotheses (Zimmerman, 2000, 2007). Although studies have examined students' scientific reasoning skills in the context of evaluating evidence, few have observed the specific scientific reasoning strategies students utilize. For instance, a common way of measuring scientific reasoning has been to use scale response items, where a given value represents the ability to reasoning more or less scientifically (Amsel et al., 2008; Klaczynski, 2001b; Macpherson & Stanovich, 2007; Sá, West, & Stanovich, 1999). This approach provides us with a general sense of how students respond to information, but the main drawback of this method is that we gain no clearer understanding about how students are reasoning. Other work has observed students' written responses (Klaczynski et al., 1997; Kosonen & Winne, 1995), which have been informative in uncovering the depth of students' scientific evaluations. Measuring scientific reasoning this way, however, only inform us of the extent to which students' scientific reasoning abilities are either poor or strong (Fong et al., 1986; Norcross et al., 1993; Zeineddin & Abd-El-Khalick, 2008). Because of these limitations, scholars have called for studies that examine that capture how students reason scientifically when evaluating evidence (Schunn & Anderson, 1999; Sodian & Bullock, 2008).

The primary goal of this study was to provide an observational account of students' scientific reasoning strategies while also assessing the depth of their scientific evaluations. By observing students' responses using these two methods, we can better understand the various ways students go about evaluating information scientifically and the extent to which their evaluations provide an informative assessment of the evidence. However, as scholars have implied, students don't always evaluate evidence scientifically (Burrage, 2008; Kuhn, 2009). Therefore, I also examined students' experiential responses, such as when students provide opinions or make judgments about the validity of evidence from their own personal experiences.

Dual process models of reasoning have been influential in explaining how people reason in a variety of tasks. This model asserts that individuals have two systems for reasoning about the world: the intuitive-experiential and the rational-analytical (Estes et al., 2003; Pacini & Epstein, 1999). The intuitive-experiential system represents reasoning that is guided by an individual's implicit intuitions, views, beliefs, and their prior experiences and expectations. The rational-analytic system, on the other hand, is guided some formal principle of logic or reasoning and is characterized as being more deliberate and effortful. In the context of evaluating evidence, I consider this rational-analytic system to be guided by principles of scientific reasoning. These two systems of reasoning do not function independently from one another, however. As Pacini and Epstein (1999) note, they can operate in an interactive parallel, where one system informs the other. Although intuitive-experiential reasoning is an optimal system for making decisions with limited time and information (Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 2002) other work has shown that people can make biased judgments from their experiences. The work on belief-bias reasoning has shown that when evidence threatens an individual's beliefs, they are more likely to view that evidence more negatively than belief-confirming evidence (Bastardi et al., in press; Lord, Ross, & Lepper, 1979; Stanovich & West, 2007; Toplak & Stanovich, 2003).

In the context of evaluating research evidence, this has implications for understanding why students may not always engage in scientific reasoning. Under the dual process model, students may be more focused on judging whether the research evidence fits with their previous beliefs and experiences than in examining the general quality of the evidence. Some line of work has found that student do not reason scientifically unless explicitly prompted to do so. It is only when prompted to be more critical that students provide more scientific evaluations (Macpherson & Stanovich, 2007; Nussbaum & Kardash, 2005). These studies, however, have not explored whether prompting students to think more critically helps decrease experiential forms of reasoning. Therefore, the second goal of this study was to examine whether providing prompts for critical thinking was influential in reducing experiential reasoning.

Based on previous work by Sá et al. (2005) and Burrage (2008), I distinguish between two types of experiential reasoning strategies, those in which students use *Opinions & Explanations* and those in which they use *Beliefs & Experiences* to evaluate evidence. Opinions & Explanations are evaluations in which students offer a personal view or explanation but do not explicitly refer to their beliefs or experiences as being a primary source of this response. Beliefs & Experiences refer to evaluations in which students are explicitly referring to their personal beliefs and prior experiences. Drawing from previous work examining students' scientific responses (Fong et al., 1986;

Klaczynski et al., 1997) I focus on three scientific reasoning strategies. These strategies include 1) *Internal Threats to Validity*, 2) *ANCOVA Reasoning*, and 3) *Methodological & Statistical* reasoning. Internal Threats to Validity captures when students evaluate studies by postulating potential problems inherent to the general design of the study, such as experimenter bias or measurement error. ANCOVA Reasoning strategies involve postulating that an observed relationship can be the result of a mediating third variable. Finally, Methodological & Statistical reasoning involves using conceptual and procedural knowledge about methods and statistical principles to evaluate research evidence, such as examining how variables were operationalized, the reasonable of the procedures, how the data was aggregated, and whether important variables were controlled for, to name a few.

Another goal of this study was to examine how differences in individuals' thinking dispositions were related to students reasoning strategies. I study two important individual thinking dispositions that may inhibit or facilitate scientific reasoning: Need for Cognition (Cacioppo & Petty, 1982) and Actively Open-minded Thinking (Stanovich & West, 1997). Cacioppo, Petty, and Morris' (1983) work on need for cognition finds that people with a high need for cognition are more motivated to evaluate information than those with a low need for cognition. Accordingly, studies have shown that people with a low need for cognition favor more experiential processing compared people with a high need for cognition, who favor more analytic approaches (Epstein et al., 1996). Stanovich's work on actively open-minded thinking has also revealed that peoples' critical reasoning is related to their willingness appreciate different viewpoints. In these studies, people with more open-minded dispositions are less likely to engage in belief-bias reasoning (Stanovich & West, 2008; Toplak & Stanovich, 2002). These studies, however, frequently use controversial or belief-threatening evidence, so it is not clear whether these thinking dispositions play a role in how students evaluate more neutral sources of evidence.

For this study, I considered how students evaluated two different types of research studies: between-group students (studies that report differences between two or more groups) and correlational studies (studies that report the association between two different variables). I focus on these two study designs because they have important implications for reasoning in everyday contexts. Findings from between-group and correlational studies are commonly reported in the media, but whether or not such findings are accurately interpreted is another matter. For instance, news outlets frequently report findings from scientific studies that were published because they had statistically significant results. But if studies have small effect sizes their findings do not necessarily warrant major changes in policy or behavior. When the effect size of a study is very small, but the findings are interpreted as large, then an "effect size" error has occurred. Although it is important to notice when small effects are being over-interpreted, and psychology journals now frequently require the inclusion of effect size information in their publications (APA, 2009), it is not clear how sensitive students are to noticing when effect size errors occur. With regards to correlational evidence, philosophers have long noted that it is impossible to draw definitive causal conclusions from correlation data and have developed a taxonomy of possible reasons why two variables A and B might be correlated even though A does not directly cause B. There are numerous claims that individuals are particularly poor at noticing when they make correlation-not-causation errors, and one of the common goals of science methodology training is to teach people to avoid such errors (Hatfield, Faunce, & Job, 2006). Although understanding the distinction between correlation and causation is important for interpreting scientific findings, we know little about whether students can notice when a correlation-not-causation error has occurred.

I created a set of study summaries that described psychological research studies, which can be found in Appendix A. These studies described social science topics and were written in a technical manner in which relevant details regarding the study design and measures were described. I had three main predictions for the results. First, I predicted that students would engage in more experiential reasoning than scientific reasoning when asked to evaluate the studies in an informal context. I also expected that, when promoted to think critically, students' would reason more scientifically while also reasoning less experientially. The depth of students' scientific evaluations would also be higher in the critical thinking condition than in the informal condition. Second, I predicted that students would be able to distinguish between studies containing interpretive errors from those that did not contain errors when prompted to critically evaluate the studies. This would be reflected by a higher number of scientific evaluations for the studies containing interpretive errors than for those without interpretive errors. Lastly, I predicted that higher Need for Cognition and Actively Open-minded Thinking scores would be positively associated with providing more scientific evaluations in both the informal and critical thinking contexts. The more a student reported a higher need for cognition and open-mindedness, the more likely they would use scientific evaluations. I also expected both dispositions to be positively associated with students' depth of scientific evaluations.

Methods

Participants

Fifty college students (38 female, 12 male) from participated in this study. Students' mean age was 18.38 years (SD = 1.02) and were freshmen or sophomores. Participants were recruited from the introductory psychology subject pool at the University of Michigan.

Procedure

This study was administered online. Once logged into the site, participants were given an hour to complete the study. Participants were asked to read and evaluate eight summaries of psychological studies. Each individual study was presented separately from one another. In addition, the studies were presented in a fixed random order. Participants were first shown these summaries in an informal context, in which I tried to elicit participants' natural responses. After reading each study, participants were asked to rate the quality of the study on a 5-point scale (1 = very poor quality… 5 = very good quality) and provide a written response to the question, "*Have these findings affected your views about [study topic]? If so, how?*"

After reviewing all the studies, participants were shown the study summaries a second time and explicitly prompted to critically evaluate the studies, "*Please critically evaluate the study (e.g., What are the strengths of the study? Is there evidence or information in the study that can be considered incorrect or misinterpreted?).*" After evaluating the study summaries in both informal and critical thinking conditions, participants completed two thinking dispositions measures, background questions, and questions that gauged their prior beliefs about the study topics. For the prior belief questions, participants were given a statement reflecting main argument of the study (e.g., for the study on children's impulsivity and grades, the statement read, "In elementary, children's academic development is a function of their ability to control their impulsive behaviors"). Using a 6-point scale (1 = disagree strongly… 6 = agree strongly), they were asked to respond to the following question: "*Prior to completing this study, the above statement would fit with my beliefs and experiences*."

Materials

Study Summaries. I created eight study summaries that provided detailed descriptions of psychological studies. These study summary topics were also considered non-belief threatening in nature (i.e., about everyday psychological topics). These summaries contained descriptive information about the study goals, sample characteristics, methodological procedures, the results, a summary describing the relevance of the study. Appendix A provides a full list of the summaries.

Study Summary Manipulations. I manipulated each study by, a) the study design, and b) the presence or absence of an interpretive error. This was done to examine how participants' evaluations changed as a function of these manipulations as well as to ensure that the studies varied from one another. For study design, half the studies contained between-group evidence and the other half contained correlational evidence. Between-group evidence reported the outcomes between two different groups (e.g., whether academic grades differ between impulsive and non-impulsive children) and correlational evidence reported the association between two different variables (e.g., whether the number of children's books in a home is associated with the number of alphabet letters children know). Half of these study summaries contained interpretive errors. For the between-group evidence, an effect size error occurred when the differences between two groups were weak but over-interpreted as having important implications. For the correlational evidence, a correlation/causation error occurred when the findings were misinterpreted as having a causal relationship. Thus, of the four study summaries with between-group evidence, two contained interpretive errors (small effect size) while the other two did not (large effect size). For the four summaries with correlational evidence, two contained interpretive errors (correlation-not-causation error) while the other two did not (correctly interpreted associative relationship).

Measures

Experiential and Scientific Evaluation Scoring. During both informal and critical thinking conditions, I examined participants' written responses for experience based and scientifically based evaluations. Experienced-based evaluations occurred when participants used their prior beliefs, experiences, or viewpoints to inform their evaluations. Scientific evaluations occurred when participants critiqued the study based on the quality and validity of the evidence.

Experiential Strategies. I used previous work by Sá et al. (2005) to distinguish between two different types of experience-based responses. The *Opinions & Explanations* code was assigned when a participant stated their views and opinions about the study or offered a personal explanation but did not explicitly refer to their beliefs, experiences as being a primary source for these responses (e.g., for the study on children's impulsivity and grades, participant 34 responded, "I do not agree with the study's conclusion and I do not think that an impatient child is bad at school because they have no control over their impulses"). The *Beliefs & Experiences* code, on the other hand, occurred when participants explicitly referred to their personal beliefs, experience, or values to evaluate the studies (e.g., participant 39 responded: "I don't think this study is strong at all because I fit the impatient stereotype and I never did poorly in school.").

Scientific Reasoning Strategies. For the scientifically reasoned response, I coded when participants employed the following three strategies: Internal Threats to Validity, ANOCVA reasoning, and Methodological & Statistical reasoning. The *Internal Threats to Validity* code captured when participants identified something inherent in the study design they believed jeopardized the validity of the study, such as experimenter and participant effects, testing errors, and faulty measures (e.g., participant 24 responded: "Due to experimenter's implicit influence on children in the experiment, children may act differently and cause experiment to be invalid."). The *ANCOVA Reasoning* code captured when participants stated the importance of some third variable that could change the results of the study (e.g., participant 10 responded: "The results of impatient children having lower grades may be compromised by parental involvement and how much priority is placed on doing well in school."). The *Methodological & Statistical Reasoning* code captured participants' methodological evaluations, such as critiquing how the participants were sampled, how the responses were quantified, the sequence of the procedures, and the strength of the findings (e.g., participant 22 responded: "The time given to the children and the bell seem a little weak.").

Depth of scientific evaluations. Depth of scientific evaluations measured the degree to which participants' scientific evaluations were reasonably explained, in terms of connecting how their evaluation was relevant for understanding the quality and the validity of the studies. Based off the work of Fong and Nisbett (1991), I used the following three-point system to code for participants' depth of scientific evaluations:

1 = *a non-scientific evaluation* captured when participants did not evaluate the study scientifically. This typically captured when participants exclusively relied on using their opinion or experiences (e.g., participant 15 responded, "This is a strong study overall, but teaching children effective self-control strategies is not the only way to improve academic achievement.").

2 = *a poor scientific evaluation* captured when participants provided a scientific evaluation, but did not explicitly state how their evaluations was relevant or important for understanding the validity of the study. This code also captured when participants' scientific evaluations were poorly defined or vague (e.g., participant 45 responded, "This study ignores extraneous variables aside from impulsive behavior, such as underlying psychological disorders).

3 = *a good scientific evaluation* captured when participants related their scientific evaluations to understanding various aspects of the study, like the quality of the measures, potential limitations in the research design, and the strength of the results (e.g., participant 43 responded, "The M&M test does not accurately test for patience. Some children could just be shy and not take the M&M, but they could be very impatient in real life.") or when participants noticed the interpretive errors for the study summaries containing flaws (e.g., The interpretation of this study is off. Children with B- averages

are not doing noticeably better than those with C+'s. The results may be significant, but it's not a large enough difference to make a meaningful impact in the real world.”).

Reasoning strategies and depth of scientific evaluation scoring procedures. For the experiential, scientific, and depth of scientific evaluation codes, composite scores were created for both the informal and critical thinking conditions. These composite scores were calculated by summing the total number of coded observations across the eight study summaries. These evaluations were not mutually exclusive and participants' responses may have included more than one type of evaluation. For example, a participant could have provided a methodological critique while also noting that the findings were consistent with their previous experiences.

Since participants evaluated the studies in two separate conditions, it was possible that participants would not evaluate the study scientifically during the critical thinking condition if they already did so during the informal condition. However, this did not turn out to be the case. Participants who provided scientific evaluations during the informal condition also did so during the critical thinking condition. Most participants provided the same scientific evaluation across both informal and critical thinking conditions. There were also instances where participants would state, “please see my previous response.” When this occurred, participants' responses in the informal condition also counted as a response for the critical thinking condition. Only in a couple of occasions did participants provide scientific evaluations in the informal condition but not in the critical thinking condition.

Two trained researchers independently coded participants' responses. For the non-directive instructions, there was a 93% agreement between both coders with Cohen's kappa value of .79. For the critical thinking instruction, there was an 89% agreement and the Cohen's kappa value was .63. Both coders resolved discrepancies through discussion. However, only one researcher coded for participants' depth of scientific evaluations.

Actively Open-minded Thinking Scale. After evaluating the summaries, participants completed the Actively Open-minded Thinking Scale (Stanovich & West, 1997). This 41-item scale assessed participants' degree of openness to different viewpoints and beliefs and flexible thinking. Twenty statements represented open-minded, flexible thinking (e.g., "I believe that the different ideas of right and wrong that people in other societies have may be valid for them") where the remaining 21 statements reflected closed-minded, rigid thinking (e.g., "No one can talk me out of something I know is right"). Participants were asked to agree or disagree with each statement using a 6-point scale (1 = disagree strongly, … 6 = agree strongly). Scores between 41 and 82 represents having closed-minded views towards other beliefs and rigid modes of thinking, scores between 83 and 123 represents having slightly narrow to slightly open-minded views, and scores above 165 represents having open-minded views towards other beliefs and flexible modes of thinking.

Need for Cognition Scale. Participants were also given the 18-item Need for Cognition Scale (Cacioppo & Petty, 1982) which measured their individual willingness to seek out, engage, and enjoy cognitively challenging activities. Ten statements reflected a high need for cognition (e.g., "I find satisfaction in deliberating hard and for long hours) while eight statements reflected a low need for cognition (e.g., "Thinking is not my idea of fun"). Participants were asked to rate the extent to which each statement was characteristic of them using a following 5-point scale (1 = extremely uncharacteristic, … 5) extremely characteristic). Scores between 18 and 36 represents having a low need for cognition, scores between 37 and 54 represents having a slightly low to slightly moderate need for cognition, and scores above 55 represent having a moderate to high need for cognition.

Results

Descriptive statistics for the outcome measures

Table 3.1 lists participants' mean quality ratings, Need for Cognition scores, Actively Open-minded Thinking scores, and reasoning strategies. Table 3.2 presents a list of the study summaries and Tables 3.2 and 3.3 provide the results of two paired *t*-tests. Regarding participants' prior beliefs, all but one study summary had average rating above 4, suggesting that the study topics fit participants' prior beliefs and experiences. None of the average prior belief ratings fell below 3 (disagree slightly) or above 5 (agree moderately), which demonstrates that the study topics were non-controversial or belief threatening in nature.

Since one of my study aims was to examine whether students noticed differences in the evidence quality (whether they contained interpretive errors), I needed to ensure that participants' prior beliefs were not drastically effecting how they evaluated the studies. In order to check for this, I used a paired *t*-test to examine whether the study summary pairs had similar prior belief ratings. As shown in Table 3.3, three of the four study summary pairs were given similar prior belief ratings. The study pair that significantly differed from one another contained the between-group, interpretive error manipulation. The study that discussed children's impulsivity and academic achievement had lower prior beliefs ratings than the study that discussed the role of social monitoring on peoples' public behaviors. Since participants' prior beliefs differed significantly for this pair, it was possible that participants would base their evaluations on whether or not they agreed with the study and not on the quality of the evidence. I used a pairwise *t*-test to examine whether these two studies were given different quality ratings. As shown in Table 3.4, participants provided similar quality ratings, suggesting that participants' prior beliefs about the study did not translate to providing different quality ratings.

Participants' quality ratings during the non-explicit instruction condition

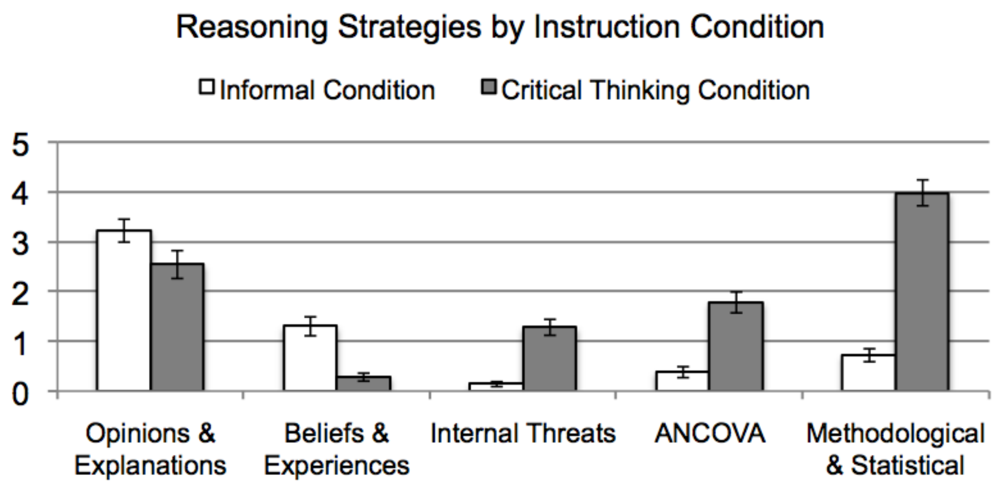
I first examined whether participants noticed when evidence contained interpretive errors during the informal condition. A 2 x 2 ANOVA using *study design* and *interpretive error* as within-subject factors revealed significant main effects. Between-group studies were given higher quality ratings ($M = 3.63$, $SD = .61$) than correlational studies ($M = 3.14$, $SD = .64$), $F(1, 49) = 19.53$, $p < .001$, $\eta_p^2 = .28$. In addition, studies containing interpretive errors were given lower quality ratings ($M = 3.08$, $SD = .63$) when compared to evidence that did not contain interpretive errors ($M = 3.69$, $SD = .55$), $F(1, 49) = 39.60$, $p < .001$, $\eta_p^2 = .44$.

There was a significant interaction between the study evidence and interpretive error manipulations, $F(1, 49) = 16.01$, $p < .001$, $\eta_p^2 = .24$. A post-hoc pairwise t-test revealed that participants gave lower quality ratings when the between-group studies contained interpretive errors (over-interpreting small effects as being large) compared to the between group studies without this error, $t(49) = 6.85$, $p < .001$. Participants' quality ratings for the correlational studies did not differ between those that contained an interpretive error (correlation-not-causation error) and those that did not.

The role of instructional prompts on participants' reasoning strategies and depth of scientific evaluations

Figure 3.1 provides the mean number of the reasoning strategies by the instructional condition.

Figure 3.1. Mean experiential and scientific reasoning strategies by instructional condition



Participants provided more opinions and explanations in the informal condition compared to the in the critical thinking condition, but this difference was marginally significant, $F(1, 49) = 3.53$, $p < .07$. However, participants did provide significantly more belief and experience based evaluations in the informal condition than in the critical thinking condition, $F(1, 49) = 22.96$, $p < .001$, $\eta_p^2 = .32$. For the scientific evaluations, participants provided significantly more internal threats to validity ($F(1, 49) = 51.34$, $p < .00$, $\eta_p^2 = .51$), ANCOVA ($F(1, 49) = 53.35$, $p < .00$, $\eta_p^2 = .52$), and methodological and statistical evaluations ($F(1, 49) = 181.30$, $p < .00$, $\eta_p^2 = .79$) in the critical thinking condition compared to the informal condition. In line with my general predictions, participants provided marginally more experiential evaluations in the informal condition. Providing prompts for critical thinking did increase participants' scientific evaluations. Interestingly however, I found that prompting students to think critically did not significantly reduce experiential reasoning, as indicated by the small decrease in their opinion and explanation-based evaluations. There was also a significant main effect of the instructional condition on participants' depth of scientific evaluations, $F(1, 49) = 216.09$, $p < .00$, $\eta_p^2 = .81$. Although participants' depth of scientific evaluations were greater in the critical thinking condition ($M = 2.05$, $SD = .44$) than the control condition ($M = .94$, $SD = .49$), both scores reflected poor scientific evaluations.

Taken together, these results suggest that participants notice whether studies contain interpretive flaws or not, as indicated by their quality ratings they provided in the informal condition. However, participants' written evaluations were mostly experiential in this context. It was only when explicitly prompted to think critically that they provided more scientifically based evaluations. Despite being prompted to critically evaluate the studies, participants still provided experientially based responses. Although participants' scientific evaluations increased in the critical thinking condition, the depth of their scientific evaluations remained generally low.

The role of study design and evidence quality on participants' reasoning strategies and depth of scientific evaluations

Scientific reasoning strategies. In this analysis, I compared whether participants were able to distinguish between studies that contained and did not contain interpretive errors and whether they used different scientific reasoning strategies according to the study design (between-group vs. correlational). Since participants did not provide many scientific evaluations in the informal condition, the analyses I present are for the critical thinking condition only. For each scientific reasoning strategy, I used a 2 x 2 ANOVA with *study design* and *interpretive error* as within-subject factors. Tables 3.5 and 3.6 contain the results for these analyses. Figures 3.2 and 3.3 below provide the results according to interpretive error for the between-group and correlational studies, respectively.

Figure 3.2. Mean scientific reasoning strategies for between-group studies

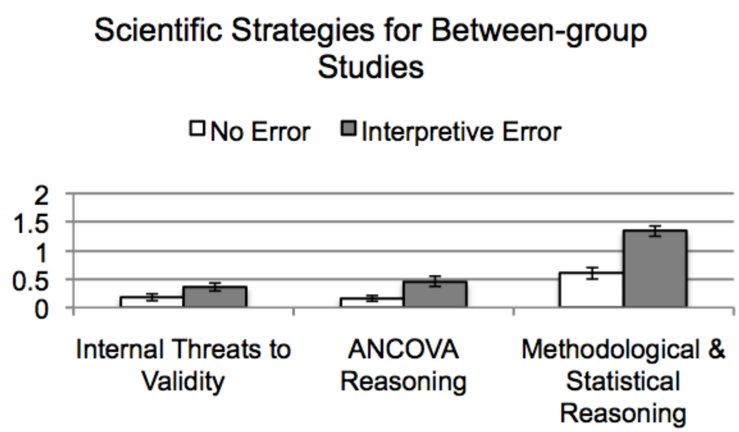
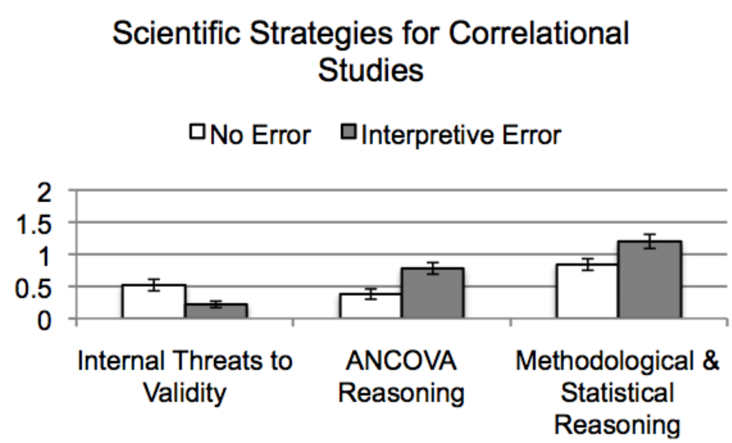


Figure 3.3. Mean scientific reasoning strategies for correlational studies

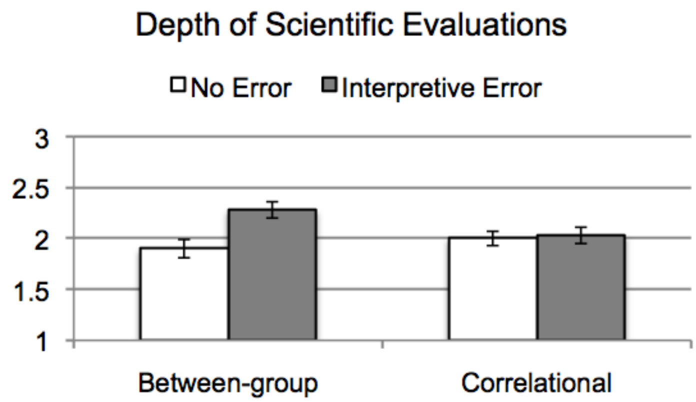


For internal threats to validity, there were no significant main effects of study design or interpretive error. However, there was a significant interaction. Participants provided more internal threats to validity when the between-group studies contained effect size errors compared to the between-group studies that did not contain this error. For the ANCOVA reasoning strategies, there was a main effect for study design. Participants provided more ANCOVA based evaluations for the correlational studies than for the between-group studies. There was also a main effect for interpretive error, in which participants provided more ANCOVA based evaluations when the studies contained interpretive errors than when they did not.

There was no main effect of study design for the methodological and statistical evaluations. There was a main effect of interpretive errors. Participants provided more methodological and statistical evaluations when the studies contained interpretive errors compared to when they did not. There was also a significant interaction, in which participants provided more methodological and statistical evaluations for the between-group studies with effect size errors compared to the between-group studies without this error.

Depth of scientific evaluations. Using the same 2 x 2 analysis, I found a main effect of interpretive error on participants' depth of scientific evaluations. Participants' depth of evaluations were higher when the studies contained flaws. There was a also a significant interaction, in which participants' depth of evaluations were higher when the between-group studies contained effect size errors compared to between-group studies that did not contain this error (See Figure 3.4).

Figure 3.4. Mean depth of scientific evaluations by interpretive error and study design manipulations



These overall results show that participants' evaluations differed according to whether or not the studies contained interpretive errors. This difference, however, was mostly attributed to participants' evaluations of the between-group studies. For these studies, participants provided less opinions and explanations while providing more scientifically based evaluations (internal threats to validity, methodological and statistical reasoning) when these studies contained weak effect sizes but were over-interpreted as being strong. With regards to the correlational studies, participants' provided more ANCOVA based evaluations for the correlational studies than for the between-group studies, which implies that they understood that third variables could explain the relationship between two correlated outcomes. They also used more ANCOVA based evaluations when the studies contained correlation-not-causation flaws that when they did not. Although participants' depth of scientific evaluations were higher for the studies containing interpretive errors, on average, these responses reflected poor scientific reasoning.

The relationship between Need for Cognition and Actively Open-minded Thinking on participants' reasoning strategies and depth of scientific evaluations

Tables 3.7 and 3.8 provide the intercorrelations between participants' individual thinking disposition scores and the reasoning outcomes for the informal and critical thinking conditions, respectively. There was a positive correlation between the actively open-minded thinking and the need for cognition measures ($r = .64$, $p < .01$). A higher disposition towards open-mindedness was associated with a higher need for cognition. However, neither thinking disposition score was significantly correlated to any of the outcomes in the informal or critical thinking conditions. There was only one marginal relationship for need for cognition in the informal condition, where a higher need for cognition was associated with providing more belief and experience-based evaluations ($r = .26$, $p < .07$).

For the critical thinking condition, the intercorrelations between the reasoning outcomes revealed some interesting relationships[1]. Providing opinion and explanation based evaluation was negatively associated with providing scientifically based evaluations. Additionally, the more a participant provided opinion and explanation based evaluations, the lower their depth of scientific evaluations ($r = -.627$, $p < .00$). This pattern suggests that participants provided either opinions and explanations or scientific evaluations, but not both at the same time. It also shows that despite being given explicit prompts for critical thinking, participants still relied on using experiential reasoning.

[1] Regarding the intercorrelations in the informal condition: Although all the experiential and scientific evaluations were positively correlated with depth of processing in this condition, it should be noted that the average depth of processing score was low, reflecting non-scientific evaluations. Thus, any response would yield a positive relationship.

Another interesting finding came from the correlations between the scientific reasoning strategies and the depth of scientific evaluation scores. The more a participant used internal threats or methodological and statistical reasoning strategies, the greater the depth of their evaluations. ANCOVA reasoning, however, was not related to the depth of scientific evaluations. This is likely due to the high number of instances in which participants simply stated the importance of a third, unobserved variable without offering any examples or explanations. For instance, the following evaluation by participant 38 reflected a common response, "I think there could be some other variables at play here."

Since participants' experiential and scientific reasoning codes were categorized by specific strategies, it was possible that distributing their evaluations in this manner hid important overarching relationships. Therefore, I added participants' opinion & explanation and belief & experience evaluations to create a composite experiential reasoning score. The same was done for participants' scientific reasoning strategies. Tables 3.9 and 3.10 provide the intercorrelations for both the informal and critical thinking conditions, respectively. These composite scores, however, did not reveal any significant relationships between Need for Cognition or Actively Open-minded Thinking.

Discussion

The goal of Study 1 was to provide an observational account of students' experiential and scientific reasoning strategies while also examining depth of their scientific evaluations. I asked students to rate the quality of study summaries, discuss the studies in a general fashion, and critically evaluate them.

In the informal condition, I found that students noticed when the studies contained interpretive errors, as reflected by the quality ratings they provided for each study. This difference was primarily due to the between-group studies. Students provided lower quality ratings when the between-group studies over-interpreted small effects compared to those that had large effects. However, when asked to provide a written response, instead of providing a scientifically based evaluation, students preferred to evaluate the studies experientially.

As expected, providing prompts for critical thinking increased scientific reasoning. The pattern of students' scientific evaluations also revealed that they recognized when studies contained interpretive errors. For the between-group studies, students provided more internal threats to validity and methodological and statistical evaluations when these studies contained effect size errors. My findings also show that students were reasonably sensitive to the limitations of correlational evidence. They provided more ANCOVA based evaluations when discussing correlational studies than when discussing between-group studies. They also provided more ANCOVA based evaluations when the correlational studies contained flaws compared to when they did not. Providing explicit prompts for critical thinking only marginally reduced their experiential responses. It seems that despite being asked to think critically, students' still preferred to evaluate the studies by discussing their personal views and experiences.

In line with my predictions, participants' depth of scientific evaluations were higher in the critical thinking condition compared to the informal condition. However, it is important to note that even though students' depth of evaluations increased, their scores remained relatively low. These findings show that there's a discrepancy between the scientific evaluations students offered and the extent to how much or how well they explained or discuss the relevance of their scientific evaluations.

The inconsistencies in the results raise some interesting questions. For instance why did students overwhelmingly prefer to reason experientially during the informal condition, despite noticing the differences in the quality of the studies? This is likely a limitation of the design of this study. Students were first asked to rate the quality of the studies using a 5-point scale and were then asked to provide a written response to the question, “Have these findings affected your views about [study topic]?” Since students evaluated the studies using the rating scale, students may have preferred to jump into discussing their personal views. Additionally, the way the question for the written was framed likely encouraged students to think about their personal views and not towards evaluating the evidence a second time.

Additionally, why didn’t prompting students to think critically significantly reduce students’ experiential reasoning and why did the depth of students’ scientific evaluations remain low in this condition? One possible reason is because the participants were composed of freshmen and sophomore students. Compared to juniors or seniors, this sample of students may have few experiences taking statistics or research methods courses, which teach students to evaluate evidence using scientific principles of reasoning. Therefore, students’ tendency to provide experiential evaluations may not entirely reflect a preference for reasoning experientially. Instead, it may simply reflect the default reasoning strategy students’ go to when uncertain about how to evaluate information scientifically.

Previous research has found links between thinking dispositions and peoples' reasoning outcomes, where a higher need for cognition and open-mindedness is positively associated with using more effort when evaluating information and avoiding biased reasoning (Cacioppo et al., 1983; Macpherson & Stanovich, 2007; Stanovich & West, 1998, 2008). However, I did not find the predicted relationship between either disposition on the observed reasoning outcomes. This may be because much of the work on individual thinking dispositions has been done under the context of evaluating controversial texts, which is thought to activate more biased forms of reasoning (Klaczynski et al., 1997). Therefore, it is possible that individual thinking dispositions do not inform how people reason when evaluating information and evidence impartial to a particular point of view. However, further research is needed to understand the relationships students' dispositions share with ability to evaluate evidence.

Finally, the intercorrelations between students' scientific reasoning and the depth of their evaluations revealed some important relationships. When students were asked to critically evaluate the studies, the more they provided internal threats and methodological and statistical evaluations, the higher their depth of their evaluations. But this wasn't the case for ANCOVA reasoning. This implies that participants do not spend much effort discussing the role of third variables when using this strategy to evaluate evidence.

Limitations & Conclusions

There are several considerations to take into account when interpreting these findings. First, the sample only consisted of freshmen and sophomore students. Compared to more advanced students, they may have the least sophisticated reasoning skills. Another issue involves the within-group manipulation. Since students had already seen the studies during the informal condition, this could have affected how they evaluated the studies during the critical thinking condition. Another limitation involved the study summaries themselves. None of the summaries contained a formal title, which would have helped students clearly identify what claims were being made from the studies. Because of this, students may have not understood the materials as easily as they would have otherwise. These summaries also contained detailed information about the study's methods, materials, designs, and controls, which may have made it demanding to comprehend. Although my intention for including detailed descriptions was so students could closely examine the measures and the general design of the study, the major drawback of providing this level of detail, however, was that it may have placed high demands on working memory, which is known to effect how well people can reason about information (De Neys, 2006). Finally, since the study was conducted online, it is unclear how engaged participants' were while taking the study.

Despite these limitations, however, this study provides a unique insight into the nature of students' critical thinking development by showing how students reason when evaluating research evidence. These observations revealed that students primarily relied on experiential systems of reasoning when thinking about evidence in an informal context. Although students noticed when studies contain interpretive flaws, they did not readily offer scientific evaluations unless explicitly instructed to do so. Additionally, even when given clear instructions for critical thinking, students still showed a preference for reasoning experientially, which may reflect a natural tendency for people to think about their views and experiences or a default strategy for reasoning people are uncertain.

By using this dual process approach, this study was effective in showing that underclassmen students use more experiential than scientific forms of reasoning as part of their repertoire for evaluating research evidence.

CHAPTER IV

Study 2

Recent work by Arum and Roska (2011) has sounded some alarms showing that students are not gaining many critical thinking skills within the first two years of college. This is especially the case regarding students' ability to critically evaluate claims and arguments. Findings from another studies have also led scholars to question whether college students' are developing the necessary scientific reasoning skills that are needed to think critically (Mill et al., 1994). Although more work is necessary to fully understand this issue, it has led educators to seriously reflect on whether college is actually teaching students important critical thinking skills (Kuhn, 2009; Kuhn et al., 2008; Leshowitz, 1989).

Despite the number of studies that have examined college students' critical thinking development, few studies have attempted to examine how students reason when evaluating claims. Traditional measures have relied on using multiple choice and Likert-type scale items to reflect whether students are thinking critically (Lehman & Nisbett, 1990; Pascarella et al., 1996; Terenzini et al., 1995). Although this provides us with a general indicator of students' critical thinking skills, it does little to reveal the process of how students arrived at a given response. Other work has used more informative measures that asks students to provide written evaluations, which are then used to examine the depth of their evaluations (Arum & Roska, 2011; Fong & Nisbett, 1991;

Klaczynski et al., 1997). This provides a better index of whether or not students are providing informed and well justified evaluations, but scoring students' responses in this way obscures the reasoning strategies students use to do so.

Dual process models of reasoning, which assumes that people utilize either intuitive-experiential systems or rational-analytic systems, provides a valuable framework for understanding the nature of students' critical thinking development. When evaluating claims and arguments, individuals can rely on their intuitive impressions, personal beliefs, or experiences to judge their validity. However, individuals may also think more rationally and analytically, in which they devote considerable effort towards evaluating the validity of a claims' supporting evidence. As prior work has demonstrated, people in general prefer to utilize intuitive-experiential reasoning when evaluating information and evidence, since it is a fast and cognitively frugal strategy for processing information (Evans, 2003; Evans & Frankish, 2009; Stanovich, 1999). This work has also shown, however, that intuitive-experiential reasoning can undermine the critical thinking process, since this form of reasoning can lead to making biased inferences from information and evidence (Amsel et al., 2008; De Neys, 2006; Evans, 1998; Evans et al., 2001). What dual process models do not fully emphasize, however, is how domain knowledge plays a role in shaping individuals' preference for using either intuitive-experiential or rational-analytic processes when evaluating evidence.

Domain-general scientific knowledge helps students discern strong from weak evidence when evaluating claims and arguments (Leshowitz et al., 2002; Schunn & Anderson, 1999). Without this knowledge, students' effortful evaluations may be misguided or underdeveloped. In examining students' critical thinking development, I argue that college helps students reason more rationally-analytically by teaching them important scientific reasoning skills. Like Study 1, the goal of this study was to provide an observational account of students' experiential reasoning (*Opinions & Explanations* and *Beliefs & Experiences*) and scientific reasoning strategies (*Internal Threats to Validity, ANCOVA Reasoning, and Methodological & Statistical Reasoning*). Unique to this study, however, is that I observe a group of freshmen and seniors majoring in psychology.

I chose to study seniors majoring in psychology because, as previous work has shown, the training social science majors receive while in college teaches them principles of scientific reasoning that are relevant for thinking critically (King et al., 1990; Lehman et al., 1988; Lehman & Nisbett, 1990; Pascarella et al., 1996; Schraw et al., 1995). For instance, psychology majors are required to take research methods and applied statistics courses, which teach them to think about how experiments are designed, how variables are operationalized and measured, and how to appropriately interpret scientific results. Students also take theoretical courses in which they learn how theories of human behavior are developed, tested, confirmed, revised, or rejected as new evidence becomes available. Because of these experiences, psychology majors will have the appropriate domain-general scientific skills that help them evaluate evidence-based claims, especially when compared to freshmen students who may enter college with little scientific training.

The second goal of my study was to understand whether anecdotal stories significantly influenced how students perceived and evaluated evidence. Some work has found that when students are presented with alluring information, such as descriptions or images of brain scans, they view psychological studies more favorably (McCabe & Castel, 2008; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). This is also the case for anecdotal stories. Prior work has shown that when an anecdotal narrative is paired alongside statistical evidence, people view the evidence more favorably (Dahlstrom, 2010; Winterbottom et al., 2008). In terms of understanding how anecdotes influence the reasoning process, when anecdotes imply that some causal relationship has occurred, this activates experiential reasoning, in which individuals quickly based their judgments on whether or not the story fits with their prior beliefs and personal experiences. My study considers whether alluring anecdotal stories influence students reduces students' ability to reason scientifically when evaluating evidence. Although seniors may be better able to reason scientifically, not much is known about whether they are able to resist being persuaded by alluring anecdotal stories.

The final goal of this study was to examine whether students' thinking dispositions and epistemic beliefs differed by class standing and whether these differences were related to students' ability to reason more scientifically. For this study, I examined two thinking disposition measures: the Rational-Experiential Inventory and the Actively Open-minded Thinking Scale. The rational-experiential inventory captures differences in students' preference for thinking either more intuitive-experientially or rational-analytically (Epstein et al., 1996). The rational-experiential inventory has been correlated with providing more in-depth scientific evaluations and avoiding belief-bias reasoning (Klaczynski et al., 1997; Pacini & Epstein, 1999). The actively open-minded thinking scale, which assesses the extent to which individuals are open to different viewpoints and have flexible modes of thinking, have also been associated with avoiding biased reasoning (Stanovich & West, 1997; Toplak & Stanovich, 2002; West et al., 2008). Although both dispositions have important implications for understanding the sources of belief-bias reasoning, not much work has explored whether exposure to college is related to students self-reported preference for thinking more rational-analytically and think more open-mindedly.

I also consider how students' epistemic beliefs differ as a function of years in college. The work on epistemic beliefs has demonstrated that college helps students think more reflectively about the nature and sources of knowledge (King & Kitchener, 1994; King et al., 1990; Kuhn, 1992). Whereas freshmen reason about knowledge in pre-reflective terms, in that knowledge is either right or wrong, college seniors understand the idea that knowledge is constructed through evidence, that knowledge can be revised or changed in light of new evidence, and that the accumulation of evidence helps establish the certainty of knowledge. For this study, I use the Epistemic Beliefs about Psychological Research scale, which assess the extent to which individuals endorse the idea that scientific findings can eventually lead establishing that a theory is certain and that one can trust the opinions of a scientific expert over one's own (Estes et al., 2003). Examining differences in student's rational-experiential dispostion, actively open-minded thinking, and epistemic beliefs can help us understand how college impacts these more reflective aspects of college students' critical thinking skills, especially with regards to students' ability to decouple their previous beliefs from examining the evidence at hand.

Unique to this study is that students are asked to evaluate articles that resemble short scientific news reports. Since these reports commonly include short anecdotal stories, it was a good platform for understanding how anecdotes influence reasoning in everyday settings. In order to assess students' reasoning outcomes, I developed a paradigm based on King & Kitcheners' (1994) Reflective Judgment Interview. Students were asked to rate the extent to which they agreed with the claims in the articles. Additionally, they were asked to provide a written response describing why they agreed or disagreed. These responses where then used to examine whether they provided intuitive-experiential or scientifically based reasons. Like Study 1, I also coded for students' depth of scientific evaluations.

I also asked students to evaluate the studies using a more established experiment evaluation scale in order to examine whether observing students' reasoning strategies correlated with this scale. Although not the main focus of my study, I also examined whether intellectual ability was correlated to students' reasoning outcomes. Prior work has demonstrated that intellectual ability is uncorrelated to the ability to avoid biased reasoning (Macpherson & Stanovich, 2007; Stanovich, 1999, 2009; Stanovich & West, 2008). Therefore, I asked students to provide their ACT scores in order to examine whether this index of intellectual ability was also unrelated to evaluating evidence-based claims.

I had three general predictions. First, I predicted that seniors majoring in psychology would demonstrate stronger critical thinking skills than would freshmen. This would be reflected by providing lower agreement ratings, providing more scientifically reasoned evaluations, having higher depth of scientific evaluation scores, and providing lower experiment evaluation ratings. Second, I predicted that seniors would be less influenced by anecdotal stories than would freshmen. Third, I predicted that students' thinking dispositions would differ by class standing, where seniors would report being more rationally-oriented, more open-minded, and have greater reflective beliefs about the nature of knowledge than freshmen. I also expected these dispositions to be associated with using less experiential reasoning, more scientific reasoning, and demonstrating less belief-bias reasoning than freshmen.

Methods

Participants

Ninety-four college students (62 female, 32 male) from the University of Michigan participated in this study. Of these students, 48 were freshmen and 46 were seniors. The majority of the participants were Caucasian (72%), followed by East Asian (12%), and African American (7%). For the freshmen participants, their mean age was 18.29 years (*SD* = .05) and reported being enrolled for a year or less at the university. Of these participants, 46 percent reported majoring or intending to major in psychology, 33 percent declared another major, and 21 percent were undecided. The seniors in the study had an average age of 21.39 years (*SD* = .95) and reported being enrolled for an average of 4.80 years (*SD* = .58). All of the seniors reported that they were psychology majors.

Most of the freshmen and all the seniors were recruited through postings advertising the study. Some of the freshmen sample was obtained from the psychology subject pool at the University of Michigan.

Procedure

The study was conducted in a lab containing desktop computers, which were separated by privacy screens. The study session was conducted with groups of four or fewer and took an hour to complete. The study was displayed using Qualtrics.com, a web-based survey site. Although the study was displayed over a web-browser, the browser was set to full-screen mode so that the study content was the only information visible to the participant. Participants were told that they would read eight articles describing psychological studies and that the goal of the study was to understand their thoughts regarding each article.

Participants were randomly assigned to either the anecdote or control condition. For the anecdote condition, each article began with an anecdotal story followed by a description of the study. The articles in the control condition only contained the description of the study. Each individual study was presented independently from one another and the Qualtrics survey program randomly determined the order of presentation.

When participants saw each article for the first time, they were asked, *"Do you agree with the claim made in the article?"* and responded using a 4-point scale (1 = strongly disagree, … 4 = strongly agree). They also provided a written response to the statement, "*Please describe why you agreed or disagreed with this claim.*" Afterwards, participants were shown the articles a second time and asked to evaluate the studies using an experiment evaluation scale (Macpherson & Stanovich, 2007). After evaluating the articles, participants completed two thinking disposition measures, an epistemic belief measure, background questions, and questions that assessed their prior beliefs and scientific knowledge. Participants who were recruited through the posting advertisements were given $15 compensation. The subject-pool participants were given course credit in exchange for participating. Since the concluding procedures differed for the recruited and subject-pool participants, both groups were studied in separate sessions.

Materials

Study Articles. I developed a set of eight articles intended to mimic short scientific news reports. The articles were written in a non-technical style and covered various psychological topics. Each article contained a title, which made a claim about the study (e.g., “Introverted Partners Decrease Marital Satisfaction”). This was then followed by a short description of the study, which included information about the sample characteristics, the sample size, procedures, measures, the results of the study, and a conclusion. Many details were intentionally left out in order to encourage skepticism towards the articles. For the anecdote condition, the article began with an anecdotal story describing scenarios intended to supported the claim being made form the described study. The control condition only contained the title of the article and the description of the study. Refer to Appendix B for the articles. Also refer to Appendices C through G for the individual measures and background questions that were used in this study.

Study Article Manipulations. For both anecdote and control conditions, the articles differed by study design (between-group vs. correlational). Four of the articles described between-group studies, which examined how two groups differed on an outcome. The other four articles described correlational studies, which examined the relationship between two different variables. Unlike Study 1, however, all of the studies contained interpretive errors. The between-group studies contained effect size errors, which the reported differences between the two groups were questionably small. For example, one of the between-group studies examining differences in siblings' creativity reported,

> As part of a recent study, researchers followed 542 sibling pairs. The oldest child, who was 10-12 years old, was given a standard creativity assessment. The younger sibling was assessed years later, the age when their older sibling had taken the assessment. The study found differences in creative abilities between siblings. Younger siblings' creativity scores were 2 points higher than older siblings' scores.

The correlational studies contained correlation-not-causation errors, in which the findings were misinterpreted as having causal relationships. The correlational study examining the relationship between social lifestyles and memory read,

> In a recent study, researchers followed 481 retired adults between 70-85 years of age. They surveyed the number of social activities they engaged in during a given week and assessed their short-term memory. The study found a positive correlation between the two variables. Having an active social lifestyle increased retired adults' short-term memory. In contrast, having an inactive social lifestyle decreased their short-term memory.

Measures

Experiential and scientific evaluation scoring. I used the same procedures as in Study 1 to code and score participants' experiential and scientific evaluations. Participants' evaluations were not mutually exclusive to one reasoning strategy, since participants often included more than one type of evaluation in their responses.

The two experiential strategies were *Opinions & Explanations* and *Belief & Experiences*. The *Opinions & Explanations* code was assigned when a participant stated their views and opinions about the study or offered a personal explanation but did not explicitly refer to their beliefs or experiences as being a primary source for these responses (e.g., regarding the article on birth order and creativity, participant 9 responded, "I do not believe there is any correlation between birth order and creativity. I feel that being born second would have nothing to do with the ability of one's creativity."). The *Beliefs & Experiences* code, on the other hand, occurred when participants explicitly referred to their personal beliefs, experience, or values (e.g., participant 87 responded, "Considering that I'm a younger sibling, I feel that younger siblings are far more creative than their older siblings. They have a great imagination and more inspiration to do creative things such as drawing, observing, and building things.")

The three scientific evaluations were *Internal Threats to Validity*, *ANCOVA Reasoning*, and *Methodological & Statistical Reasoning*. The *Internal Threats to Validity* code captured when participants identified something inherent in the study design they believed jeopardized the validity of the study, such as experimenter and participant effects, testing errors, and faulty measures (e.g., participant 63 responded, "…Also, this is a poor assessment of identifying creativity because the subjects are siblings. The older who took the test beforehand could have told the younger sibling about the test which could account for the younger sibling doing better on the test."). The *ANCOVA Reasoning* code captured when participants stated the importance of some third variable that could change the results of the study (e.g., participant 17 responded, "There could be other factors at play here, such as perhaps the parents give the younger child more attention than the older child, and that attention could be the cause of differences in creative ability, not just their birth order."). The *Methodological & Statistical Reasoning* code captured participants' methodological evaluations, such as critiquing how the participants were sampled, how the responses were quantified, the sequence of the procedures, and the strength of the findings (e.g., participant 72 responded, "If the study were to be more valid, and the claim more believable, the experiment should have tested each sibling at the same time, not at the same age. When the older sibling took the assessment, so should have the younger sibling on the same day.").

Depth of scientific evaluations. This was similar to Study 1 in that I assessed the degree to which participants' scientific evaluations were reasonably explained and sensitive to noticing interpretive errors. However, since all of the articles in this study contained interpretive errors, I modified my scoring to capture when participants' identified these errors, which is similar to that of Fong and Nisbetts' (1991) and Klaczynski and colleagues' (1997) schemes. I used a four-point system to code participants' depth of scientific evaluations.

1 = a *non-scientific evaluation* captured when participants did not evaluate the study scientifically and instead used their opinions and explanations or belief and experiences. For the article reporting on active social lifestyles and memory, participant 33 responded, "I agree with the study because the more you use your mind the stronger it remains. Your mind is like any other muscle. The more you use it, the better you will be."

2 = a *poor scientific evaluation* captured when participants provided a scientific evaluation or statement but did not clearly explain how their evaluations were relevant or important for understanding the quality or the validity of the study. Participant 82 responded, "There may have been other variables to account for their better short-term memory."

3 = a *good scientific evaluation* captured when participants attempted to relate their scientific evaluations to understanding the quality and validity of the study.

Participant 38 responded, "…I wouldn't say that being social alone increases memory - there are a lot of factors that need to be addressed in this study, especially illnesses that target cognitive processes in the elderly. For example someone with Alzheimer's probably has a lower working memory and less social interaction as well."

4 = a *strong scientific evaluation* captured when participants noticed the interpretive errors in the study. Participant 89 responded, "I disagree because just because the two variables are positively correlated does not mean that socializing causes better memory. It could equally be asserted that having a better memory causes a person to socialize more." For the article regarding siblings and creativity, participant 29 responded, "Though it may have a small effect, the test scores increased only by 2 points and this can be accredited to chance and not necessarily to birth order."

Reasoning strategies and depth of scientific evaluation scoring procedures. A composite score was calculated by summing the total number of coded observations across the eight study articles. The depth of scientific evaluation score was the averaged score across all the articles or across the articles with similar study designs (between-group vs. correlational). Two trained researchers independently coded participants' reasoning strategies and depth of scientific evaluation scores.

Experiment Evaluation Scale. After rating whether they agreed with the articles and providing written justifications, participants were asked to re-read and evaluate the articles using an experiment evaluation scale (Macpherson & Stanovich, 2007). This scale asked participants to evaluate the study summaries on three dimensions: the strength of the claim, the persuasiveness of the study, and overall quality of the study. These questions read as follows: *How strongly is the claim supported by the results of* *the study?; How persuasive was this study?; What is your overall evaluation of the quality of this study?* Participants responded to these questions using a 6-point scale, where 1 represented an unfavorable evaluation and 6 represented a favorable one. I obtained a composite experiment evaluation score by summing the score for the three questions for each summary. Scores between 3 and 6 represented unfavorable evaluations, scores between 9-12 represented slightly unfavorable to slightly favorable evaluations, and scores above 13 represented favorable evaluations.

Actively Open-minded Thinking Scale. This 41-item scale assessed participants' degree of openness to different views and beliefs and flexible thinking (Stanovich & West, 1997). Responses are on a 6-point scale, ranging from "disagree strongly" to "agree strongly." Scores between 41 and 82 represents having closed-minded views towards other beliefs and rigid modes of thinking, scores between 83 and 123 represents having slightly narrow to slightly open-minded views, and scores above 165 represents having open-minded views towards other beliefs and flexible modes of thinking.

Rational-Experiential Inventory. The 40-item Rational-Experiential Inventory (Epstein et al., 1996) measured participants' preference for thinking using either intuitive-experiential or rational-analytic systems. Eighteen statements reflect a preference for thinking analytically (e.g., Using logic usually works well for me in figuring out problems in my life.), whereas twenty-two statements reflect a preference for thinking intuitively (e.g., I hardly ever go wrong when I listen to my deepest gut feelings to find an answer). Participants were asked to agree or disagree with the statements using the following 5-point scale, ranging from "strongly disagree" to "strongly agree." Scores between 40 and 80 represents having a strong to moderate preference towards intuitive-experiential thinking, between 81 and 120 represents having a slight preference for intuitive-experiential to a slight preference for rational-analytical thinking, and scores above 121 represents having a moderate to strong preference for rational-analytical thinking.

Epistemic Beliefs About Psychological Research Scale (EBPR) (Estes et al., 2003). This scale asks participants' to agree or disagree with 4 statements regarding one's certainty about social scientific knowledge and one's confidence in scientific experts. In Estes and colleagues' (2003) study, participants were told that the statements referred to research on children's psychological health and development. However, since the articles in this study discussed a wide range of age groups, the word 'children' was taken out so the instructions read, "*The following statements refer to research on psychological health and development*." Responses were on a 7-point scale, ranging from "disagree very strongly" to "agree very strongly." Two epistemic beliefs sub-scores were obtained by summing the two items regarding the certainty of social scientific knowledge and the two items regarding confidence towards scientific experts. Participants scoring between 4 and 5 disagreed with the view that scientific knowledge could be certain and that scientific experts' opinions could be trusted over one's own opinion. Those scoring between 6 and 9 were uncertain. Participants scoring above 10 agreed with the view that scientific knowledge could be certain and that scientific experts' opinions could be trusted over their own.

Prior beliefs. Participants were shown the title for each article (e.g., "Active Social Lifestyles Improves Retired Adults' Memory") and asked to respond to the statement, "*Prior to completing this study, the above claim would've fit with my beliefs and experiences*." Participants' rated their level of agreement on a 4-point scale, ranging from "strongly disagree" to "strongly agree."

SAT and ACT scores. Participants provided self-reports of their SAT and/or ACT composite scores. If participants didn't fully remember their scores, they were asked to estimate what they think they received. The SAT scores were not used in the analysis because these responses were noticeably varied, in that participants provided different combinations of their math, witting, critical reading, multiple choice, and essay scores.

Scientific knowledge. I assessed participants' knowledge about the scientific method using five questions. Participants were asked, "*Are you familiar with the general principles of the scientific method?*" Two additional questions asked if they were familiar with the principles regarding between-group and correlational designs. At the end of the survey, participants were also asked where they were familiar about effect size and correlation-not-causation errors. These questions read, "*Are you familiar with the idea that one must examine the size of the quantitative research effect (a.k.a. effect size) in order to determine how large or important the result is*?," and "*Are you familiar with the idea that just because two variables are correlated doesn't mean that one causes the other*?" Participants' rated their knowledge confidence on a 5-point scale, ranging from "not at all familiar" to "very familiar." I created a scientific knowledge score by averaging the scores across the five question items.

Results

Descriptive statistics for the outcome measures

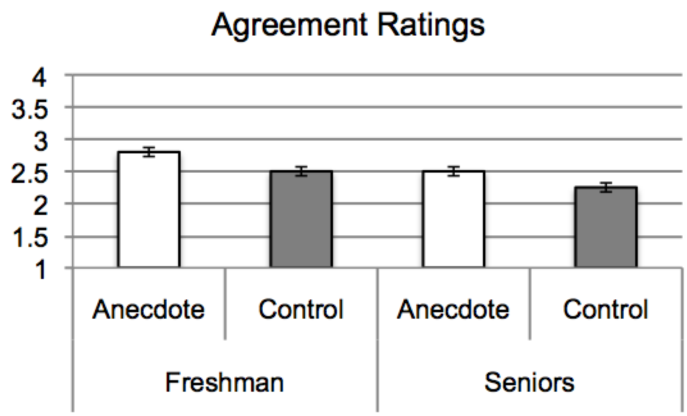
Table 4.1 list participants' individual differences measures and their evaluative outcomes. In regards to their overall evaluations, participants neither agreed nor disagreed with the claims in the article (M = 2.52, SD = .40). Also, compared to other reasoning strategies, participants preferred to use opinion and explanation based reasoning when describing why they agreed or disagreed with the claims. These descriptive results also show that participants did not use many scientific evaluations.

Regarding the low instances of belief and experience based responses, it is important to note that this was primarily the result of how this response was coded. This code was only assigned when participants explicitly referred to some belief or experience. It is just as likely that participants' opinion and explanation were based on personal beliefs or experiences. However, participants' opinions and explanations were not informative enough to confirm this with certainty (e.g., regarding the article on birth order and creativity, participant 10 responded, "I agree with this statement because birth order has a lot to do with who someone might turn out to be.").

The role of anecdotal information on freshmen and seniors' agreement ratings

For this analysis, I used a 2 x 2 x 2 ANOVA with *anecdotal information* and *class standing* as between-subjects factors and *study design* as a within subject factor (*sex* was included as a covariate). Tables 4.2 and 4.3 provides the mean agreement ratings and the analyses for the three factors, respectively. Below, Figure 4.1 presents the mean agreement ratings.

Figure 4.1. Mean agreement ratings by class standing and anecdote condition



I found a significant main effect for anecdotal information, $F(1, 89) = 11.78$, $p < .001$, $\eta_p^2 = .12$. Participants in the anecdote condition agreed more with the claims in the article than participants in the control. There was also a main effect for class standing, where seniors agreed less with the claims in the article than freshmen, $F(1, 89) = 18.35$, $p < .000$, $\eta_p^2 = .17$. There was no significant interaction between the anecdotal information and class standing. Instead, a linear relationship was found. Although the anecdotal information influenced participants' agreement ratings, freshmen agreed more with the claims than seniors. The same trend was found for the control condition; participants in the control condition agreed less with the claims, but seniors agreed even less so than freshmen. When examining the study design factor, both freshmen and seniors' agreement ratings did not differ for the articles containing either between-group or correlational designs.

The role of anecdotal information on freshmen and seniors' reasoning strategies and depth of scientific evaluations

Reasoning Strategies. For these analyses, I used the same 2 x 2 x 2 design to examine participants' experiential and scientific reasoning strategies. The means for each reasoning strategy are presented in Table 4.4 and analyses for the three factors are presented in Table 4.5. Figures 4.2 and 4.3 presents the mean reasoning strategies by the anecdote condition for the freshmen and seniors, respectively.

Figure 4.2. Freshmen students' reasoning strategies by anecdote condition

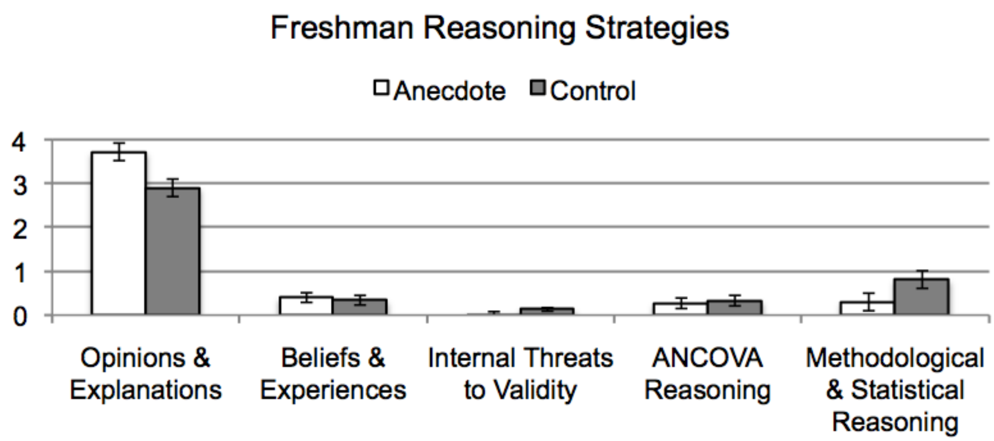
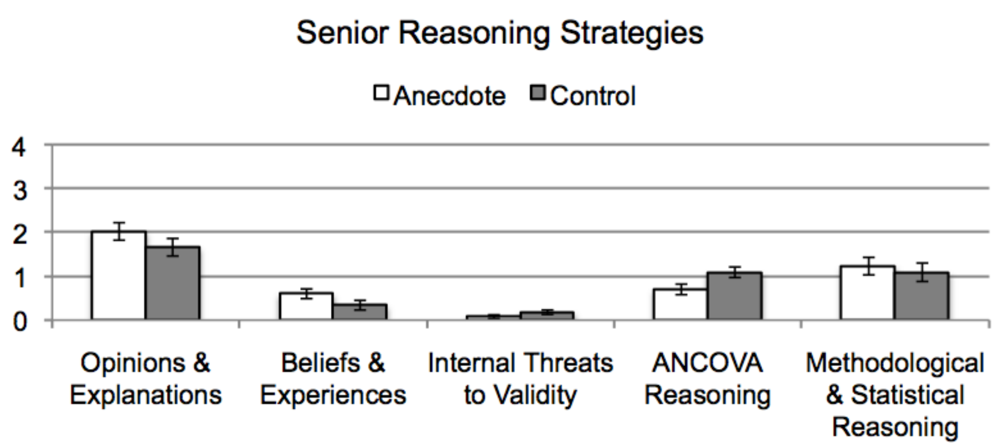


Figure 4.3. Senior students' reasoning strategies by anecdote condition



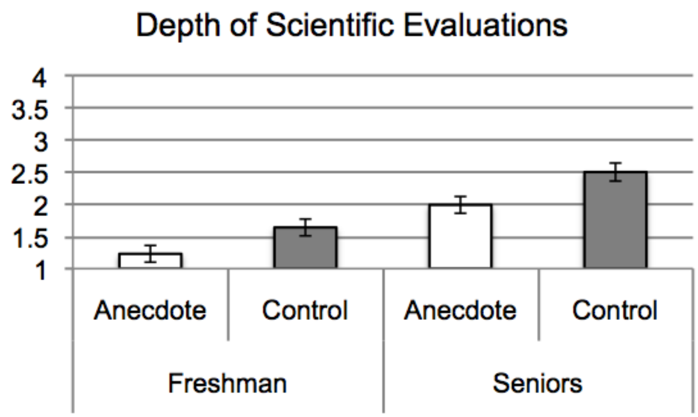
For the main effects, participants in the anecdote condition gave marginally more opinion and explanation-based responses than participants in the control condition, $F(1, 89) = 3.52$, $p < .06$, η_p^2. There was no main effect for the beliefs and experience-based responses. The results for the scientific reasoning strategies supported my general predictions. Although the number of internal threats to validity participants provided were virtually non-existent, participants in the anecdote condition provided significantly more internal threats to validity than participants in the control condition, $F(1, 89) = 4.17$, $p < .04$, $\eta_p^2 = .05$. This was also the case for ANCOVA reasoning, although these results were marginally significant, $F(1, 89) = 3.79$, $p < .055$, $\eta_p^2 = .04$. Participants in the anecdote condition also used more methodological and statistical reasoning than participants in the control, $F(1, 89) = 8.37$, $p <.005$, $\eta_p^2 = .09$.

For class standing, seniors provided significantly less opinions and explanations than freshmen, $F(1, 89) = 39.88$, $p < .0001$, $\eta_p^2 = .31$, and provided more methodological and statistical evaluations compared to freshmen, $F(1, 89) = 25.37$, $p < .0001$, $\eta_p^2 = .22$. There were no interactions between anecdotal information and class standing. Like before, this relationship was linear, in which anecdotal information influenced freshmen's reasoning strategies more than it influenced seniors' reasoning strategies. Finally, there were no main effects of study design on any of the reasoning strategies. Participants provided the same reasoned responses for the articles containing between-group and correlational designs.

Depth of scientific evaluations. I used the same 2 x 2 x 2 design to examine participants' depth of scientific evaluations. The means of depth of scientific evaluation scores are presented in Table 4.6 and analyses for the three factors are presented in Table

4.7. Below, Figure 4.4 shows participants depth of scientific evaluations by the anecdote condition and class standing.

Figure 4.4. Mean depth of scientific evaluations by class standing and anecdote condition

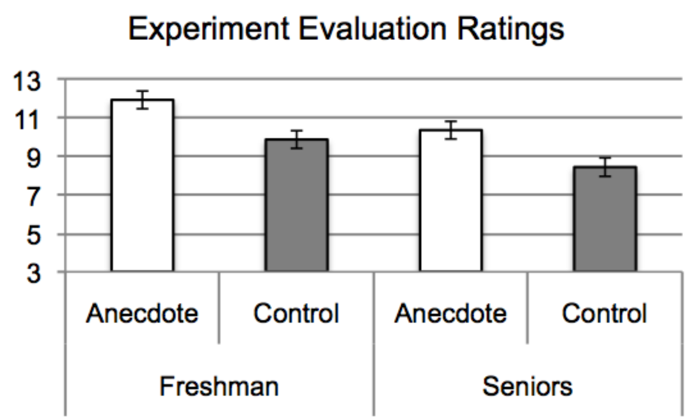


As expected, participants in the anecdote condition had a lower depth of scientific evaluations than participants in the control condition, $F(1, 89) = 10.60$, $p < .002$, $\eta_p^2 = .11$. Seniors' depth of scientific evaluations were higher than freshmen, $F(1, 89) = 34.37$, $p < .0001$, $\eta_p^2 = .28$. Although there was no significant interaction, the same linear trend was found. For the anecdote condition, seniors had a higher depth of scientific evaluations score than freshmen. Although the depth of evaluations was greater in the control condition, seniors showed even greater depth of evaluations than freshmen. There were no significant main effects for study design. Participants' depth of scientific evaluations was similar for articles containing between-group and correlational designs.

The role of anecdotal information on freshmen and seniors' experiment evaluation ratings

Using the same 2 x 2 x 2 analyses as before, I found that the main effects of anecdotes supported my predictions. The mean experiment evaluation ratings are presented in Table 4.8 and analyses for the three factors are presented in Table 4.9. Participants in the anecdote condition provided significantly more favorable experiment evaluation ratings than participants in the control, $F(1, 89) = 18.28$, $p < .0001$, $\eta_p^2 = 17$. There was also a main effect of class standing, where seniors' experiment evaluation ratings were less favorable than freshmen's' experiment evaluation ratings, $F(1, 89) = 10.49$, $p < .002$, $\eta_p^2 = .11$. There was no significant interaction (See Figure 4.5).

Figure 4.5. Mean experiment evaluation ratings by class standing and anecdote condition



With regards to the study design, across both anecdote and control conditions, participants gave more favorable experiment evaluation ratings to articles containing correlational studies ($M = 10.52$, $SD = 3.03$) than articles containing between-group studies ($M = 9.81$, $SD = 2.60$), $F(1, 89) = 3.97$, $p < .05$, $\eta_p^2 = .04$.

Differences between freshmen and seniors' ACT scores, thinking dispositions, epistemic beliefs, and scientific knowledge

Table 4.10 displays the pairwise *t*-tests examining whether freshmen and seniors differed in their individual characteristics.

Table 4.10. Paired t-test of freshmen and seniors' individual difference measures

Variable	Freshmen	Seniors	paired *t*-test
ACT Score (n = 83)	28.50 (2.83)	27.80 (3.68)	.98
REI	126.46 (13.17)	125.09 (11.97)	.53
AOT	178.46 (17.17)	186.08 (17.36)	-2.23*
EB Certainty	9.33 (1.83)	8.96 (1.81)	1.00
EB Confidence	11.27 (1.62)	10.98 (1.91)	.80
Scientific Knowledge	3.52 (.64)	4.35 (.50)	-6.92***

*$p < .05$, ***$p < .001$

There were no differences in ACT scores between freshmen and seniors. For the thinking disposition measures, I predicted that seniors would report having higher actively open-minded thinking that freshmen. This prediction was confirmed. Although both groups rated themselves as having moderate open-minded views, seniors reported a greater preference for open-mindedness than freshmen. However, freshmen and seniors' rational-experiential scores were not significantly different; both groups rated themselves as having a moderate preference for thinking rationally. I also predicted that seniors would have more reflective epistemic beliefs about scientific knowledge than freshmen, but these results were insignificant. Both freshmen and seniors somewhat agreed with the idea that scientific knowledge could eventually be certain and that they would trust the opinions of scientific authorities over their own. Regarding participants' scientific knowledge, seniors reported having a greater familiarity with the principles of the scientific method than freshmen.

Intercorrelations between participants' individual difference measures and their reasoning outcomes

For the following intercorrelations, I created a composite experiential reasoning score, which was the sum of the opinions & explanations and the beliefs & experiences scores. I also created a composite scientific reasoning score by summing the three scientific reasoning strategy scores. Table 4.11 presents the intercorrelations between the individual differences measures (ACT scores, rational-experiential ratings, open-mindedness ratings, epistemic beliefs ratings, prior beliefs, scientific knowledge) and five evaluative outcomes (agreement ratings, experiential reasoning, scientific reasoning, depth of processing scores, experiment evaluation ratings). The eleven participants who did not provide ACT scores (4 freshmen and 7 seniors) were excluded from this analysis.

Intercorrelations between the individual difference measures. The ACT scores were not related to the thinking disposition or epistemic belief scores. Participants' intellectual ability, as indicated by the ACT score, was not related to other individual measures. There were no significant relationships between the rational-experiential inventory, the actively open-minded thinking scale, or the epistemic beliefs scales. There was positive relationship with the rational-experiential inventory and scientific knowledge, where participants who reported a greater preference for thinking rationally also reported having greater scientific knowledge.

The actively open-minded thinking scale was positively related to the epistemic beliefs scale assessing confidence in scientific authorities. Participants who reported being more open-minded were more likely to agree that they would trust the opinions of a scientific expert over their own opinions. Actively open-minded thinking was also positively related to scientific knowledge.

Relationship between the individual difference measures and reasoning outcomes. ACT scores were not associated with any of the reasoning outcome. These findings support previous research showing the ability to evaluate evidence is independent of cognitive ability. In line with my predictions, I found that actively open-minded thinking was significantly related to participants' reasoning outcomes. Participants who reported being more open-minded provided less experiential reasoning ($r = -.27$, $p < .01$), more scientific reasoning ($r = .27$, $p < .01$), had greater depth of processing ($r = .30$, $p < .01$), and provided lower experiment evaluation ratings ($r = -.23$, $p < .05$). I also predicted the same relationship for the rational-experiential inventory. However, I found no relationships for this individual measure on the observed reasoning outcomes. The results for the epistemic belief measures were mixed. These results were either non-significant or were related in the opposite direction from what I expected. For instance, regarding participants' certainty of scientific knowledge, the more a participant agreed with the idea the scientific evidence could eventually be certain, the more favorable experiment evaluation ratings they provided ($r = .23$, $p < .05$).

Prior beliefs were significantly correlated with all of the evaluative outcomes. Participants who stated that the claims in the article fit their beliefs and experiences gave higher agreement ratings, provided more experiential reasoning, less scientific reasoning, had lower depth of scientific evaluations, and provided more favorable experiment evaluation ratings. As predicted, scientific knowledge was significantly correlated with all of the evaluate outcomes. Participants with greater scientific knowledge gave lower agreement ratings, provided less experiential reasoning, more scientific reasoning, had higher depth of processing, and provided less favorable experiment evaluation ratings.

Interestingly, actively open-minded thinking was not related to prior beliefs. This may imply that individuals with high levels of open-mindedness are better able to decouple their prior beliefs when evaluating evidence.

Lastly, participants' reasoned evaluations were highly correlated with the experiment evaluation scores. The more a participant provided an experiential evaluation, the higher their experiment evaluation rating ($r = .57$, $p < .01$). Conversely, the more a participant provided a scientific evaluation the lower their experiment evaluation ratings ($r = -.60$, $p < .01$). These correlations suggest that assessing participants' reasoning strategies is a good indicator of critical thinking. Furthermore, the intercorrelations between the reasoning strategies and the experiment evaluation scale revealed how participants used either strategy; participants who provided favorable evaluations tended to reason more experientially whereas participants who provided unfavorable evaluations reasoned more scientifically.

Discussion

The goal of this study was to provide a cross-sectional account of college students' critical thinking development. I compared freshmen students with seniors majoring in psychology to determine whether there were differences in students' critical thinking abilities. Most of my predictions were confirmed. When examining students reasoned responses, seniors agreed less with the articles than freshmen. Seniors also provided more scientific evaluations and less experiential evaluations when describing why they agreed or disagreed with the articles when compared to freshmen. The depth of seniors' scientific evaluations were higher than freshmen students' depth of evaluations. Finally, seniors provided lower experiment evaluations ratings than freshmen.

Together, these findings demonstrate that senior psychology majors have stronger critical thinking skills than freshmen students. Although this study is cross-sectional, these findings have guiding implications for understanding how college potentially influences how students reason in everyday situations. As my results show, college seniors reported having greater knowledge about the scientific method than freshmen, and differences in students' scientific knowledge were related to how often and how well they reasoned scientifically. It is possible that taking courses relevant to understanding the scientific process helps students learn important domain-general scientific reasoning strategies.

Since seniors were better at reasoning scientifically than freshmen, I also expected them to be better able to resist alluring anecdotal information when evaluating evidence-based claims. Therefore, I predicted that seniors would be better able to think critically when faced with alluring anecdotal stories than freshmen. This prediction was partially supported. Seniors in the anecdote condition agreed less with the articles, provided more scientific evaluations, and gave lower experiment evaluation scores than freshmen in the anecdote condition. However, the main effect of anecdotes was consistent across class standing in that both freshmen and seniors evaluated the articles more favorably than the freshmen and seniors in the control condition. So although seniors showed stronger critical thinking skills than freshmen, they still were susceptible to being influenced by alluring anecdotal information. The depth of seniors' scientific evaluations also indicates this. In the anecdote condition, seniors' scientific evaluations were generally poor (M =

1.98, *SD* = .72), whereas seniors in the control condition had more in-depth scientific evaluations (*M* = 2.46, *SD* = .87).

My findings also revealed that freshmen and seniors had different levels of open-mindedness, where seniors reported being more open-minded to different viewpoints than freshmen students. However, I did not find my predicted results for the rational-experiential inventory and the two epistemic beliefs about psychology research scores. Although freshmen reported having the same preference for thinking rationally as seniors reported, freshmen overwhelmingly preferred to reason experientially about the claims. This finding shows that freshmen students may not have accurate perceptions of how they believe they prefer to think and how they actually think when evaluating evidence.

I was also surprised to find no differences between freshmen and seniors' epistemic beliefs. These non-significant findings go against previous work, which shows that students enter college with a tendency to reason pre-reflectively about the nature and sources of knowledge. In this study, freshmen reported having the same beliefs about scientific research as seniors. Although I don't have a clear explanation for these findings, it may be that the epistemic beliefs about psychological research scale does not adequately capture more varied aspects of how students view the nature and sources of knowledge. This scale only measures whether students believe scientific research can be certain and whether they would trust a scientific experts' opinion. In order to get a more accurate measure of epistemic beliefs, it is important to capture various dimensions relevant to how students' understand the nature and sources of knowledge. For instance, based on Hofer & Pintrichs' (2002) conceptions of epistemic beliefs, additional questions could assess how students understand what counts as knowledge, whether knowledge is simple or complex, where knowledge comes from, how knowledge is justified, and the extent to which different sources of knowledge are certain.

Interestingly, I found that actively open-minded thinking was related to students reasoning outcomes in very interesting ways. Students' level of open-mindedness was not related to their prior beliefs ($r = -.09$). However, open-mindedness was associated with agreeing less with the articles, providing less experiential evaluations, and providing more scientific evaluations. Additionally, open-mindedness was associated with providing more in-depth scientific evaluations and providing lower experiment evaluations ratings. These solid relationships across all of the reasoning outcomes suggest that that open-minded thinking bears an important relationship with the ability to avoid biased reasoning and with providing in-depth scientific evaluations. These correlations, however, are partially confounded by the fact that more seniors reported being more open-minded than freshmen. Therefore, it can also be the case that seniors in general are better able to avoid biased reasoning that freshmen students. In order to address this issue, future studies should focus on obtaining a greater sample of seniors and examining whether within-group differences in their level of open-mindedness is related to avoiding belief-bias reasoning.

Limitations and conclusion

This study showed that, when given evidence-based claims to evaluate, senior psychology majoring were better able to think critically than freshmen. This may largely be because, as previous work has shown, social science students may be better trained to think about these types of ill-structured tasks when compared to students who major in math or business, or the natural sciences, for example (Burrage, 2008; King et al., 1990; Lehmann, 1963; Norcross et al., 1993). Therefore, these findings should not be generalized beyond understanding the development of critical thinking skills among psychology majors. One potential limitation of this study is that it is uncertain whether the anecdotes were influencing individuals by activating experiential responses, since I only found a marginal relationship between students' experiential evaluations and the anecdote condition manipulation. The length of the articles may also potentially explain why students viewed articles containing anecdotal stories more favorably. Prior work has found longer articles are judged as being better written than shorter ones (Petty & Cacioppo, 1986; Pierro, Mannetti, Erb, Spiegel, & Kruglanski, 2005). Therefore, students in the anecdote condition may have been less critical because they thought the articles were better written – as opposed to being persuaded to think that the anecdotal stories provided conclusive, experience-based support for the claims. However, Study 2b address this concern.

These results have implications for understanding what students, especially those majoring in psychology, learn in college. These findings imply that senior psychology majors are learning important scientific reasoning skills that help them approach critical thinking more purposefully than freshmen. These seniors also seem to learn how to think more open-mindedly in college, which is closely related to the ability to decouple one's prior beliefs from the reasoning process. Although college potentially helps students' shift away from using less experiential forms of reasoning and more towards rational-analytic ones (or scientific, in this case), seniors still have difficulty resisting persuasive anecdotal information when evaluating claims.

CHAPTER V

Study 3

An alternative explanation for the results of Study 2a is that participants in the anecdote condition were simply influenced by the length of the articles and not the anecdotal stories themselves. As previous work has shown, people think longer articles are better written than shorter ones (Petty & Cacioppo, 1986; Pierro et al., 2005). This may have influenced students in the anecdote condition to evaluate the articles based on how well they were written and not on the persuasiveness of the anecdotal stories. Therefore, I conducted a follow-up study to see whether the length of the articles was influencing students to believe that they were better written. In order to control the effects of participants' prior beliefs, they were reminded to separate their beliefs from their evaluations throughout the study.

Methods

Participants, procedures, and measures

Thirty-six college students (12 females, 24 males) with a mean age of 18.94 years (*SD* = 1.14) participated in this study. All of the students were recruited from the psychology subject-pool. This sample was composed of 72% freshmen, 17% sophomores, 3% juniors, and 8% seniors.

Participants came into the lab and were presented with the same materials as before. In this instance, however, they were told that the articles were presented for possible publication in a local news site and that we wanted their feedback regarding how well the articles were written. Eighteen participants were randomly placed in the anecdote condition and eighteen in the control.

Writing quality rating. When each article was presented, they were given the following instructions, "Regardless of how you feel about the claims being made in this article, please base your responses on the quality of the writing." Based on measured used by Cacioppo, Petty, & Morris (1983), they were asked to judge the quality of the writing on three dimensions using a 6-point scale. The first statement read, "*Considering both content and style, please rate how well this article was written*," and the responses ranged from "very poorly written" to "very well written." The second statement read, "*Please rate how well a person will be able to understand this article*," and responses ranged from "very difficulty to understand" to "very easy to understand." The final statement read, "*Please rate the structure of the article*," and the responses ranged from "contains very complex structure" to "contains very simple structure." I created a writing quality composite score by summing the three responses. Scores between 3 and 6 were rated as being a poorly written article, scores between 9 and 12 were neither poorly or well written, and scores above 13 were rated as being a well-written article.

Experiment evaluation scale. After rating the quality of the writing, participants saw the articles a second time. In this instance, they were given the following instructions, "Regardless about how you feel about the claims being made in this article, please base your responses on the quality of the study presented in this article," and rated the article using an experiment evaluation scale (Macpherson & Stanovich, 2007). Afterwards, participants completed the same thinking disposition and background questionnaires.

Results and Discussion

Writing quality. I used a 2 x 2 ANOVA with *anecdotal information* as a between-subject factor and *study design* as within-subject factor on the writing quality score. I found no main effect for anecdotal information on the quality of the writing, $F(1, 34) < 1$, *ns*. The quality of the writing was rated similarly for the anecdote ($M = 13.63$, $SE = .32$) and control conditions ($M = 13.56$, $SE = .32$). There were also no differences for the study design, $F(1, 34) < 1$, *ns*. Articles containing between-group studies were given the same writing quality ratings ($M = 13.54$, $SE = .26$) as articles containing correlational studies ($M = 13.64$, $SE = .24$).

Experiment evaluation scale. Using the same 2 x 2 ANOVA analysis, I found no main effect of anecdotal information on the experiment evaluation ratings, $F(1, 34) < 1$, *ns*. Participants' experiment evaluation ratings were similar in the anecdote ($M = 10.78$, $SE = .47$) and control conditions ($M = 10.43$, $SE = .47$). There was a main effect for study design, where participants provided lower ratings when the articles contained between-group studies ($M = 9.92$, $SE = .45$) compared to the articles containing correlational studies ($M = 11.30$, $SE = .32$), $F(1, 34) = 11.56$, $p < .002$, $\eta_p^2 = .24$.

These results demonstrate that, when asked to decouple their beliefs when evaluating the articles, the anecdotal stories did not influence students to think that they were better written than the articles in the control condition. This provides some support for my argument that anecdotal stories elicit experiential reasoning, which in turn influence students to perceive and evaluate the articles more favorably.

Intercorrelations between individual difference measures, the writing quality ratings, and experiment evaluation ratings. Since this study asked participants to separate their prior beliefs from their responses, I did not expect to find any associations between participants' prior belief scores and their evaluations. However, this did not turn out to be the case. As shown in Table 5.1, there was a positive association between prior beliefs and the writing quality ratings. Participants who were more likely to state that the articles fit with their beliefs and experiences were also more likely to rate the quality of the writing as being higher. The same association was found between participants' prior beliefs and their experiment evaluations scores. Although asking participants to separate their beliefs when evaluating the articles counteracted the influence of anecdotal stories, students still had trouble separating their beliefs and experiences when evaluating the claims.

Interestingly, there were significant associations between ACT scores across the individual difference measures. However, ACT was only marginally related to the writing quality rating and unrelated to the experiment evaluation rating. The rational-experiential inventory was positively related to actively open-minded thinking. Although the results for the ACT and rational-experiential inventory scores are different from the findings in Study 2, it is possible that these relationships were caused by the nature of this study. Participants were explicitly reminded to decouple their experiences from their evaluations throughout the study. It is therefore possible that when taking the thinking disposition surveys, participants may have based their answers according to whether or they thought they were avoiding biased responding. However, whether or not this reflects a more accurate portrayal of participants' thinking dispositions is difficult to say, since it is uncertain whether the task influenced students to respond more honestly or more extremely towards the thinking disposition items.

CHAPTER VI

General Discussion

Using a dual process framework, which assumes that individuals rely on using quick, intuitive-experiential responses more so than deliberate, rational-analytic responses, the goal of my dissertation was to understand whether college training helped students reason less experientially and more scientifically when evaluating research evidence. My dissertation also advanced how researchers examine college students' critical thinking skills by observing the reasoning processes students used when evaluate evidence.

Across both studies I found that students preferred to use more experiential reasoning strategies than scientific ones when evaluating evidence. Study 1 showed that college underclassmen overwhelmingly preferred to reason experientially when evaluating research studies in an informal context. Although prompting students to think critically increased scientific reasoning, it did not significantly decrease their experiential evaluations. The overall descriptive statistics for Study 2 also found that students preferred to provide more experientially based evaluations than scientific ones when asked to describe why they agreed or disagreed with a set of claims that were made from news-like science articles. Both findings support dual process views of reasoning, in that students approached thinking critically using more autonomous, experiential systems than deliberate, rational ones.

Do these results show that students fail to reason scientifically when critically evaluating evidence? Not necessarily. As discussed in Study 1, even college underclassmen were able to notice when research evidence contained interpretive errors when asked to rate the quality of the studies. On possible reason there was a discrepancy between students' quality ratings and their preference for providing experiential responses could be due to the design of the task. In the informal condition, students were asked to rate the quality of the studies on 5-point scale and provide a written response to the question, "Have these findings affected your views about [study topic]?" Because the quality ratings had an evaluative component, students may have simply preferred to jump into discussing their personal views and experiences when providing a written response, having already made a judgment about the quality of the evidence. Additionally, since the question asked students to think about their views, this may have influenced students to reason more experientially than they would have otherwise. Therefore, the results from students' responses in the informal condition are somewhat confounded by how the questions were ordered and framed.

The findings from the critical thinking condition provide more room for interpretation, however. Prompting students to think critically increased scientific reasoning. Students also provided more scientific evaluations when the studies contained interpretive errors. Interestingly, their scientific evaluations were also strategic. For instance, they provided more ANCOVA evaluations for the correlational studies than for the between-group studies. They also provide more methodological and statistical evaluations when the between-group studies contained errors compared to the between-group studies that did not contain this error. However, the depth of students' scientific evaluations were relatively low. Additionally, students still gave a high number of experiential evaluations in this context. This may imply that students had difficulty articulating their scientific evaluations and may have therefore provided experiential evaluations as a sort of default strategy for reasoning. Thus, the results of Study 1 show that even underclassmen could think critically about evidence, but their written scientific evaluations did not reflect a strong scientific understanding. This lack of evaluative depth may have been due to students' general lack of scientific knowledge.

Building off of these findings, the primary goal of Study 2 was to examine the relationship between domain-general scientific knowledge and students' ability to evaluate evidence scientifically. More simply put, I wanted to know whether college training helped students acquire adequate scientific knowledge, and whether this helped students reason more scientifically than experientially when evaluating evidence. I compared seniors majoring in psychology (who I assumed acquired some relevant domain-general scientific strategies from their statistics and research methods coursework) with freshmen students. Students read a set of articles that made claims about various social-scientific studies. These studies, however, contained clear interpretive errors, such as over-interpreting small effects and implying causation from correlations findings.

When students were asked to rate how much they agreed with the claims, seniors disagreed more with the claims when compared to freshmen. Additionally, when asked to describe why they agreed or disagreed with the claims, seniors provided more scientific evaluations and less experiential evaluations than freshmen students. Seniors' evaluations also reflected a greater ability to reason strategically; when the studies contained effect size errors, seniors provided more methodological and statistical evaluations, and when the studies contained correlation-not-causation errors, they provided more ANCOVA based evaluations. Finally, seniors' scientific evaluations were more in-depth than freshmen students' evaluations.

Although I can't imply causation from these results, these comparative findings provide a guiding framework for understanding how college training potentially helps students shift away from using experiential systems of reasoning and more towards scientific ones. Seniors reported having significantly greater scientific knowledge than freshmen, which provides some support for the view that college training, specifically in psychology, helps students develop important domain-general scientific skills.

The seniors in this study demonstrated the ability to think critically, in that they were less willing to agree with claims containing flawed interpretations and provided more scientifically reasoned evaluations. However, another important critical thinking skill involves being able to resist persuasive information. Scientific news reports often begin with a short narrative account about a person or a scenario that helps highlight the relevance of a study. Although anecdotal narratives are useful for facilitating comprehension and learning, they can also undermine critical thinking. Anecdotal stories are believed to activate experiential systems of reasoning, in which individuals make judgments about claims based on whether the story fits with one's own experiences or whether the story seems believable. Therefore, including anecdotal narratives alongside scientific news reports may inadvertently influence students to think more about their beliefs and experiences than the general quality of the scientific evidence.

Study 2 additionally compared whether senior psychology majors were also able to resist the persuasive allure of anecdotal narratives. Half of the students in this study received articles that contained short anecdotal stories followed by a description of the study. For the anecdote condition, seniors were less influenced by the anecdotal stories than freshmen. However, seniors in the anecdote condition still agreed more with the claims and provided more experiential evaluations when compared to the seniors in the control condition. These findings show that despite demonstrating stronger scientific reasoning skills than freshmen, senior psychology majors still had difficulty ignoring anecdotal stories. These findings lend support to the view that anecdotal narratives are influential in activating experiential systems of reasoning.

The final goal of my dissertation was to examine important reflective aspects of the critical thinking process. I examined whether differences in thinking dispositions, such as need for cognition, actively open-minded thinking, and rational-experiential thinking, were related to how students critically evaluated evidence. I additionally examined the role of epistemic beliefs on students' critical thinking. In Study 1, I expected two important thinking dispositions (need for cognition, actively open-minded thinking) to be negatively associated with providing experiential evaluations. Conversely, I expected these thinking dispositions to be positively associated with providing scientific evaluations. However, I failed to find these expected relationships. This may be partly due to the characteristics of the study summaries. The summaries students evaluated was written using technical language and provided detailed descriptions of the methods. The content of these summaries contrasts previous thinking dispositions studies, in which individuals are asked to evaluate evidence that either support or threaten their previous beliefs. In studies where individuals' beliefs are brought to bare, the need for cognition and actively open-minded thinking have been associated with providing less biased evaluations. Therefore, the materials for Study 1 were not well suited for capturing the expected relationships.

Study 2 also examined the relationship between actively open-minded thinking on students' reasoning outcomes. In addition, I used the rational-experiential inventory to examine whether the disposition to think more rationally (versus experientially) was related to students' reasoning outcomes. Finally, using the epistemic beliefs about psychology research scale, I wanted to understand whether viewing psychological research as a valid source of knowledge was related to students' reasoning outcomes.

When comparing these thinking dispositions by class standing, I found that seniors reported being more open-minded towards different viewpoints than freshmen. Whether through general maturation, exposure to the college environment, coursework, or through the culmination of all these factors, the senior psychology majors in this sample seemed to have developed a greater appreciation for different perspectives and greater openness towards belief change. Although my findings are not developmental, they coincide with previous work showing that college students become more open-minded as they grow older.

Open-minded thinking was correlated to the reasoning outcomes in a way to suggest that this disposition is important for understanding the critical thinking process. Students who reported being more open-minded agreed less with the claims in the articles, provided less experiential evaluations, and provided more scientific evaluations.

Additionally, students' open-mindedness was positively related to having more in-depth scientific evaluations.

These relationships were interesting, considering that students' ACT scores were not significantly related to any of the reasoning outcomes. The non-significant findings from the ACT scores lend support to Stanovich's claim that reflective aspects of individuals' thinking dispositions, like level of open-mindedness, are more important for being able to avoid biased reasoning than intellectual ability. However, it should be noted that seniors and freshmen differed on actively open-minded thinking. Therefore, the correlational results for the actively open-minded thinking scores are confounded by the fact that seniors reported being more open-minded than freshmen.

Different from my expectations, I did not find any significant relationships between students' rational-experiential inventory scores, their epistemic beliefs about psychological research scores, and their reasoning outcomes. This was especially puzzling for the rational-experiential inventory, a measure that has previously been associated with the ability to reason scientifically. The epistemic beliefs about psychology research scale, on the other hand, was limited in that it only captured whether students viewed psychological research with certainty and whether they would change their views if they differed from that of a scientific expert.

Based on my dissertation findings, I present a revision of my initial model in Figure 6.1, which represents college students' critical thinking development. Different from my previous model, I excluded the 'college experience' box in favor of a 'college training' box, which represents important school-related factors that help develop students' critical thinking skills (college major, courses taken relevant to understanding scientific principles). This in turn has an influence on students' domain-general scientific knowledge, which is now represented in the 'individual factors' box. This was done in order to emphasize the important relationship I found between scientific knowledge and students' reasoning outcomes. Also, my previous model assumed that students enter college with little scientific knowledge, and that college experience is the cardinal factor that helps students learn to think scientifically. But it is not possible to assume that this is the case. Students' scientific knowledge may have also been shaped by their prior experiences in high school or elsewhere. Additionally, as the results of Study 1 showed, even underclassmen, who may not have experience taking courses relevant to the scientific method, could think scientifically in informal contexts.

My findings also revealed that actively open-minded thinking had an important relationship with students' reasoning outcomes. Therefore, my individual factors box highlights this construct. Although I did not find relationships between students' epistemic beliefs and their reasoning outcomes, I still included this disposition in model, given that the measures I used did not adequately capture this construct. Finally, I placed the individual factors box as a moderator for students' reasoning systems, given that students' preference for thinking experientially or scientifically may initially be guided these factors.

Aside from my findings, my studies have important implications for instruction. Scholars have emphasized that helping students think more critically involves teaching both deliberate skills, such as the ability to reason scientifically, and reflective ones, like the ability to decouple one's beliefs from the critical thinking process. What is less emphasized, however, is the importance of teaching students to be able to identify when claims and arguments contain alluring information, since such information can potentially undermine the critical thinking process. One effective way to helps students' resist alluring information would be to assign activities in which students identify instances in the media, whether on television or in newsprint, where anecdotal stories are used to support some claim or argument. Such activities would help students' apply their critical thinking skills towards evaluating evidence in an everyday context.

My studies have also helped advanced how critical thinking is measured. Observing students' reasoned evaluations uncovered the various ways they utilized specific scientific reasoning strategies, such as providing ANCOVA based evaluations for correlational studies or providing methodological and statistical evaluations when the between-group studies contained effect size errors. By observing students' non-scientific evaluations, I was also able to show that students' often discussed their personal views or opinions when evaluating research evidence. The findings I obtained from this observational approach also supported the underlying assumption of dual process models, in that people often favor intuitive-experiential forms of reasoning over rational-analytic ones. Finally, this approach was effective in capturing the differences in freshmen and senior psychology majors' evaluations. Since students' experientially- and scientifically-based evaluations correlated strongly with students' experiment evaluation scale ratings (a more traditional, Likert-type measure of critical thinking), this shows that this observational approach is effective in capturing how students critically evaluate research evidence.

In order to understand how college training influences students' reasoning systems, their ability to evaluate evidence-based claims, and their thinking dispositions, longitudinal work is needed that examines these constructs over the course of students' time in college. It would also be beneficial to further understand how college majors differentially promote particular rational-analytic skills and specific dispositions towards thinking.

Table 3.1. Descriptive statistics for the quality ratings, thinking dispositions, reasoning strategies, and depth of processing

Variable (N = 50)	*M*	*SD*	Range
Quality Ratings	3.39	0.49	2.40
Need for Cognition	58.40	10.01	44.00
Actively Open-minded Thinking	166.76	19.66	85.00
Informal Condition			
Opinions & Explanations	3.22	1.66	7.00
Beliefs & Experiences	1.30	1.36	5.00
Internal Threats	0.14	0.35	1.00
ANCOVA Reasoning	0.38	0.76	4.00
Meth & Stats Reasoning	0.72	0.95	4.00
Depth of Processing	0.94	0.50	1.75
Critical Thinking Condition			
Opinions & Explanations	2.54	1.96	8.00
Beliefs & Experiences	0.28	0.57	2.00
Internal Threats	1.28	1.16	5.00
ANCOVA Reasoning	1.78	1.52	5.00
Meth & Stats Reasoning	3.98	1.82	7.00
Depth of Processing	2.05	0.44	2.00

Table 3.2. List of study summary topics and prior beliefs paired t-test results

Study Summary Topic	Study Design	Error
A. Children's impulsivity and grades	between-group	interpretive error
B. Public monitoring and social obedience	between-group	interpretive error
A. Children's aggression learning	between-group	no error
B. Presumptive questions and memory	between-group	no error
A. Home environment and letter knowledge	correlational	interpretive error
B. Motivation and career achievement	correlational	interpretive error
A. Children's task frustration and anger	correlational	no error
B. Social engagement and memory	correlational	no error

Table 3.3. Paired t-test of prior beliefs scores (standard deviations in parentheses)

Study Summary Manipulations	Summary A	Summary B	paired *t*-test, within groups
Between-group			
Interpretive error	3.34 (1.17)	4.88 (1.00)	-7.32***
No error	4.40 (1.03)	4.40 (1.34)	.00
Correlational			
Interpretive error	4.28 (1.39)	4.42 (1.20)	-.57
No error	4.22 (1.00)	4.16 (0.98)	.30

***$p < .001$

Table 3.4. Paired t-test of quality rating scores (standard deviations in parentheses)

Study Summary Manipulations	Summary A	Summary B	paired *t*-test, within groups
Between-group			
Interpretive error	3.02 (1.27)	3.24 (1.17)	-.93
No error	4.04 (1.00)	4.22 (0.93)	-1.00
Correlational			
Interpretive error	3.14 (1.21)	2.92 (1.05)	1.04
No error	3.18 (1.10)	3.32 (1.10)	-.60

Table 3.5. Mean reasoning strategies and depth of scientific evaluations by study design and interpretive error (standard deviations in parentheses)

	Study Design			
	Between-group		Correlational	
	Error	No Error	Error	No Error
Opinions & Explanations	.42 (.64)	1.00 (.86)	.58 (.73)	.54 (.65)
Beliefs & Experiences	.16 (.37)	.04 (.20)	.04 (.20)	.04 (.20)
Internal Threats to Validity	.36 (.53)	.18 (.44)	.22 (.42)	.52 (.70)
ANCOVA Reasoning	.46 (.65)	.16 (.37)	.78 (.68)	.38 (.60)
Method & Stats Reasoning	1.34 (.66)	.60 (.80)	1.20 (.78)	.84 (.68)
Depth of Scientific Eval	2.28 (.60)	1.90 (.65)	2.03 (.61)	2.00 (.51)

Table 3.6. Results of reasoning strategies and depth of processing by study design and interpretive error

	F	*p*	h_p^2
Opinions & Explanations			
Study Design	3.13	.08	--
Interpretive Error	12.66	.00	.21
Study Design x Interpretive Error	9.83	.00	.17
Beliefs & Experiences			
Study Design	3.13	.08	--
Interpretive Error	3.13	.08	--
Study Design x Interpretive Error	3.13	.08	--
Internal Threats to Validity			
Study Design	2.60	.12	--
Interpretive Error	.70	.41	--
Study Design x Interpretive Error	9.33	.00	.16
ANCOVA Reasoning			
Study Design	14.8	.00	.23
Interpretive Error	23.77	.00	.33
Study Design x Interpretive Error	.42	.52	--
Method & Stats Reasoning			
Study Design	.27	.61	--
Interpretive Error	39.80	.00	.45
Study Design x Interpretive Error	4.34	.04	.08
Depth of Scientific Evaluations			
Study Design	1.56	.22	--
Interpretive Error	10.37	.00	.18
Study Design x Interpretive Error	7.51	.01	.13

Table 3.7. Intercorrelations for thinking dispositions, quality ratings, prior beliefs, reasoning strategies, and depth of scientific evaluations in the informal condition

Variable	1	2	3	4	5	6	7	8	9
1. NFC	--								
2. AOT	.64**	--							
3. Quality Ratings	.09	.12	--						
4. Prior Beliefs Ratings	.05	.06	.23	--					
5. Opinion & Explanations	-.11	-.04	.25	-.04	--				
6. Belief & Experiences	.26†	.06	.14	.14	.07	--			
7. Internal Threats	.00	.04	-.07	.11	-.12	.34*	--		
8. ANCOVA Reasoning	.10	.14	.18	.15	.08	.15	-.05	--	
9. Meth & Stats Reasoning	-.20	-.14	-.23	-.26†	-.05	.07	.24†	-.02	--
10. Depth of Scientific Eval	-.05	-.10	.02	-.08	.34*	.53**	.52**	.33*	.63**

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking.

$^{\dagger}p < .10$, $*p < .05$, $**p < .01$

Table 3.8. Intercorrelations for thinking dispositions, quality ratings, prior beliefs, reasoning strategies, and depth of scientific evaluations in the critical thinking condition

Variable	1	2	3	4	5	6	7	8	9
1. NFC	--								
2. AOT	.64**	--							
3. Quality Ratings	.10	.12	--						
4. Prior Beliefs Ratings	.05	.06	.23	--					
5. Opinion & Explanations	.00	.06	-.28*	.05	--				
6. Belief & Experiences	.00	.20	-.08	-.01	.28*	--			
7. Internal Threats	-.10	-.04	.33*	.13	-.54**	-.15	--		
8. ANCOVA Reasoning	.13	.15	.12	-.04	-.24†	-.09	.30*	--	
9. Meth & Stats Reasoning	-.16	.00	-.02	.01	-.35*	-.07	.30*	.15	--
10. Depth of Scientific Eval	-.80	-.08	.16	.02	-.63**	-.18	.48**	.20	.65**

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking.

† $p < .10$, * $p < .05$, ** $p < .01$

Table 3.9. Intercorrelations for thinking dispositions, quality ratings, prior beliefs, composite reasoning strategy scores, and depth of scientific evaluations in the informal condition

Variable	1	2	3	4	5	6
1. NFC	--					
2. AOT	.64**	--				
3. Quality Ratings	.09	.12	--			
4. Prior Belief Ratings	.05	.06	.23	--		
5. Experiential Composite Score	.08	.00	.28†	.06	--	
6. Scientific Composite Score	-.08	.00	-.08	-.70	.12	--
7. Depth of Scientific Eval	-.05	-.09	.02	-.08	.58**	.78**

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking.

†$p < .10$, *$p < .05$, **$p < .01$

Table 3.10. Intercorrelations for thinking dispositions, quality ratings, prior beliefs, composite reasoning strategy scores, and depth of scientific evaluations in the critical thinking condition

Variable	1	2	3	4	5	6
1. NFC	--					
2. AOT	.64**	--				
3. Quality Ratings	.09	.12	--			
4. Prior Belief Ratings	.05	.06	.23	--		
5. Experiential Composite Score	.00	.10	-.27†	.04	--	
6. Scientific Composite Score	-.06	.06	.16	.03	-.50**	--
7. Depth of Scientific Eval	-.08	-.08	.17	.02	-.61**	.64**

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking.

†p < .10, *p < .05, **p < .01

Table 4.1. Descriptive statistics for the individual differences measures and evaluative outcomes

Variable (N = 94)	*M*	*SD*	Range
ACT Scores (n = 83)	28.17	3.26	15.00
REI	125.79	12.55	71.00
AOT	182.03	17.63	84.00
EB Certainty	9.15	1.82	8.00
EB Confidence	11.13	1.77	8.00
Agreement Ratings	2.52	0.4	2.00
Reasoning Strategies			
Opinions & Explanations	4.96	2.27	8.00
Belief & Experiences	0.84	1.06	5.00
Internal Threats	0.21	0.44	2.00
ANCOVA Reasoning	1.17	1.28	6.00
Meth & Stats Reasoning	2.07	2.23	8.00
Depth of Scientific Evaluations	1.82	0.78	2.75
Experiment Evaluation Ratings	10.16	2.52	11.38

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking: EB Certainty = epistemic beliefs about the certainty of scientific knowledge: EB Confidence = epistemic beliefs about one's confidence in scientific experts

Table 4.2. Mean agreement ratings by anecdote condition, study design, and class standing with sex as a covariate (standard deviations in parentheses)

	Anecdotal Information Condition					
	Anecdote			Control		
	Overall	Between-group	Correlational	Overall	Between-group	Correlational
Freshmen	2.80 (.28)	2.68 (.30)	2.91 (.45)	2.53 (.37)	2.46 (.44)	2.59 (.45)
Seniors	2.47 (.27)	2.42 (.37)	2.52 (.35)	2.24 (.44)	2.28 (.42)	2.20 (.59)

Table 4.3. Results of agreement ratings by anecdote condition, study design, and class standing with sex as a covariate

Variable	*F*	*p*	h_p^2
Anecdote Condition	11.77	.001	0.12
Study Design	0.01	.940	--
Class Standing	18.35	.000	0.17
Anecdote x Study Design	1.66	.200	--
Study Design x Class Standing	2.69	.110	--
Anecdote x Class Standing	0.08	.780	--

Table 4.4. Mean reasoning strategies by anecdote condition, study design, and class standing with sex as a covariate (standard deviations in parentheses)

	Anecdotal Information Condition					
	Anecdote			Control		
	Overall	Between-group	Correlational	Overall	Between-group	Correlational
Opinions & Explanations						
Freshmen	6.54 (1.47)	3.08 (.82)	3.45 (.93)	5.80 (1.77)	2.70 (1.08)	3.08 (.92)
Seniors	4.04 (1.89)	1.75 (1.11)	2.29 (1.23)	3.31 (2.37)	1.68 (1.12)	1.63 (1.46)
Beliefs & Experiences						
Freshmen	.79 (1.21)	.45 (.65)	.33 (.70)	.67 (.92)	.45 (.65)	.20 (.50)
Seniors	1.20 (1.31)	.87 (1.03)	.33 (.48)	.68 (.57)	.31 (.47)	.36 (.58)
Internal Threats						
Freshmen	.08 (.28)	.00 (.00)	.08 (.28)	.25 (.44)	.20 (.41)	.04 (.20)
Seniors	.16 (.48)	.00 (.00)	.16 (.48)	.36 (.49)	.36 (.49)	.00 (.00)
ANCOVA Reasoning						
Freshmen	.54 (.72)	.33 (.56)	.20 (.41)	.67 (1.00)	.29 (.55)	.37 (.57)
Seniors	1.37 (1.09)	.58 (.65)	.79 (.97)	2.18 (1.56)	.95 (.84)	1.22 (1.02)
Meth & Stats Reasoning						
Freshmen	.58 (.97)	.33 (.76)	.25 (.60)	1.63 (1.88)	.91 (1.01)	.70 (1.00)
Seniors	2.46 (2.10)	1.54 (1.35)	.91 (1.10)	3.77 (2.49)	2.13 (1.52)	1.63 (1.39)

Table 4.5. Results of reasoning strategies by anecdote condition, study design, and class standing with sex as a covariate

Variable	*F*	*p*	h_p^2
Opinions & Explanations			
Anecdote Condition	3.51	.060	.04
Study Design	.07	.790	--
Class Standing	39.88	.000	.30
Anecdote x Study Design	1.18	.220	--
Study Design x Class Standing	.31	.570	--
Anecdote x Class Standing	.00	.970	--
Beliefs & Experiences			
Anecdote Condition	1.95	.160	--
Study Design	1.50	.220	--
Class Standing	.96	.170	--
Anecdote x Study Design	1.99	.160	--
Study Design x Class Standing	.14	.710	--
Anecdote x Class Standing	.90	.350	--
Internal Threats			
Anecdote Condition	4.16	.040	.05
Study Design	1.32	.250	--
Class Standing	1.20	.270	--
Anecdote x Study Design	19.75	.000	.18
Study Design x Class Standing	.37	.550	--
Anecdote x Class Standing	.03	.870	--
ANCOVA Reasoning			
Anecdote Condition	3.79	.055	.04
Study Design	1.32	.250	--
Class Standing	25.41	.000	.22
Anecdote x Study Design	.72	.400	--
Study Design x Class Standing	1.90	.170	--
Anecdote x Class Standing	2.16	.150	--
Meth & Stats Reasoning			
Anecdote Condition	8.38	.005	.08
Study Design	.04	.850	--
Class Standing	25.37	.000	.22
Anecdote x Study Design	.01	.920	--
Study Design x Class Standing	2.80	.100	--
Anecdote x Class Standing	.12	.730	--

Table 4.6. Mean depth of scientific evaluation scores by anecdote condition, study design, and class standing with sex as a covariate (standard deviations in parentheses)

	Anecdotal Information Condition					
	Anecdote			Control		
	Overall	Between-group	Correlational	Overall	Between-group	Correlational
Freshmen	1.22 (.26)	1.26 (.35)	1.18 (.58)	1.65 (.61)	1.68 (.68)	1.60 (.67)
Seniors	1.98 (.72)	2.08 (.81)	1.88 (.94)	2.46 (.87)	2.59 (1.01)	2.32 (.95)

Table 4.7. Results of depth of scientific evaluation scores by anecdote condition, study design, and class standing with sex as a covariate

Variable	*F*	*p*	h_p^2
Anecdote Condition	10.60	.002	.11
Study Design	0.01	.950	--
Class Standing	34.37	.000	.28
Anecdote x Study Design	0.08	.780	--
Study Design x Class Standing	0.88	.350	--
Anecdote x Class Standing	0.05	.830	--

Table 4.8. Mean experiment evaluation ratings by anecdote condition, study design, and class standing with sex as a covariate (standard deviations in parentheses)

	Anecdotal Information Condition					
	Anecdote			Control		
	Overall	Between-group	Correlational	Overall	Between-group	Correlational
Freshmen	11.92 (2.17)	11.60 (2.30)	12.23 (2.86)	9.84 (2.52)	9.66 (2.70)	10.03 (2.84)
Seniors	10.35 (1.92)	9.76 (2.02)	10.93 (2.57)	8.42 (2.27)	8.09 (2.22)	8.75 (2.94)

Table 4.9. Results of experiment evaluation ratings by anecdote condition, study design, and class standing with sex as a covariate

Variable	*F*	*p*	h_p^2
Anecdote Condition	18.28	.000	.17
Study Design	3.97	.050	.04
Class Standing	10.48	.002	.11
Anecdote x Study Design	0.27	.600	--
Study Design x Class Standing	0.60	.440	--
Anecdote x Class Standing	0.02	.882	--

Table 4.11. Intercorrelations for ACT scores, thinking dispositions, epistemic beliefs, prior beliefs, scientific knowledge, composite reasoning strategy scores, depth of processing, and experiment evaluation ratings

Variable (n = 83)	1	2	3	4	5	6	7	8	9	10	11
1. ACT Score	--										
2. REI	.11	--									
3. AOT	.05	.15	--								
4. EB Certainty	-.14	-.08	-.06	--							
5. EB Confidence	.04	.04	.28**	.18†	--						
6. Prior Belief Ratings	.04	-.03	-.09	.10	.04	--					
7. Scientific Knowledge	.04	.26*	.32**	-.11	-.01	-.06	--				
8. Agreement Ratings	-.01	-.12	-.18†	.18†	-.12	.47**	-.44**	--			
9. Experiential Composite Score	.15	-.07	-.27**	.04	-.08	.25*	-.46**	.64**	--		
10. Scientific Composite Score	.13	.07	.27**	-.07	.04	-.34**	.46**	-.68**	-.90**	--	
11. Depth of Scientific Eval	.14	.10	.30**	-.10	.05	-.32**	.44**	-.65**	-.85**	.95**	--
12. Experiment Eval Rating	-.06	-.10	-.23*	.23*	-.03	.44**	-.43**	.74**	.57**	-.60**	-.59**

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking: EB Certainty = epistemic beliefs about the certainty of scientific knowledge: EB Confidence = epistemic beliefs about one's confidence in scientific experts

$^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .01$

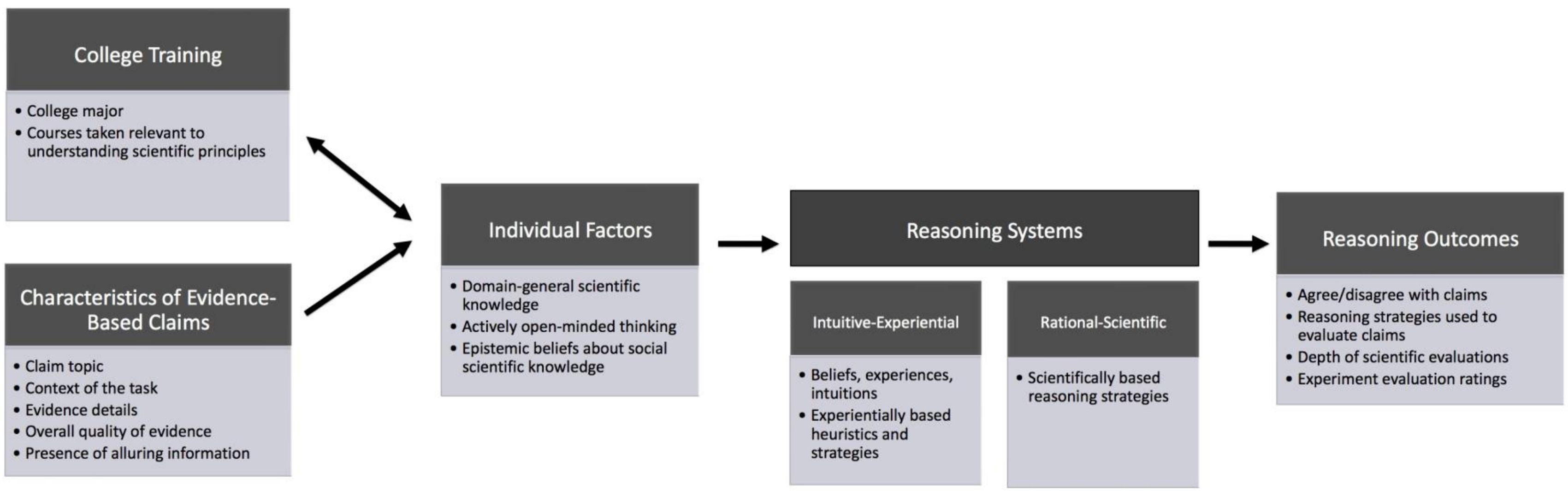
Table 5.1. Intercorrelations for ACT scores, thinking dispositions, epistemic beliefs, prior beliefs, scientific knowledge, writing quality ratings, and experiment evaluation ratings

Variable	1	2	3	4	5	6	7	8
1. ACT Score	--							
2. REI	.06	--						
3. AOT	.35†	.39**	--					
4. EB Certainty	.33†	.34*	.28†	--				
5. EB Confidence	-.02	.13	.34*	.19	--			
6. Prior Belief Ratings	.33†	-.35*	-.09	-.16	-.18	--		
7. Scientific Knowledge	.38*	.09	.09	.34*	.05	.33*	--	
8. Writing Quality Rating	.30†	.27	.26	.24	.26	.41*	.27	--
9. Experiment Eval Rating	-.02	-.04	-.04	-.37*	.05	.34*	.15	0.3†

Note: NFC = Need for Cognition: AOT = Actively Open-minded Thinking: EB Certainty = epistemic beliefs about the certainty of scientific knowledge: EB Confidence = epistemic beliefs about one's confidence in scientific experts

$^{\dagger}p < .10$, $*p < .05$, $**p < .01$

Figure 6.1. Revised model of college students' critical thinking development



Appendices

Appendix A. Study 1 Summaries

Note: The interpretive error manipulations (no interpretive error, interpretive error) are highlighted. These sections were not highlighted for participants.

Children's Aggression learning (between-group study with no interpretive error)

A group of psychologist wanted to examine the relationship between watching aggressive acts and aggressive behavior in children. In this study, 4 and 5 year old children were randomly assigned to two different conditions. In the first condition, children watched a 5-minute video of two men, Rocky and Johnny, play with a set of toys. In the video, Johnny does not want Rocky to play with his toys, so Rocky uses aggression (pushing and kicking) to take the toys away from Johnny. In the second condition, children watched a video of Johnny and Rocky playing cooperatively (sharing toys). After children watched the video, they were left alone in a playroom for 20 minutes. The experimenters observed the children through a one-way mirror and recorded their behaviors. The experimenters found significant differences between the two conditions.

Children who watched the aggressive video engaged in 38 individual acts of aggressive behavior (throwing, kicking, punching toys), whereas children who watched the non-aggressive video engaged in 11 individual acts of aggressive behaviors.

This study suggests that children can be induced to behave aggressively by showing videos of people engaging in aggressive acts. These findings raise some serious concerns regarding the extent to which violent programming on television encourages aggression in children.

Presumptions questions and false memories (between-group study with no interpretive error)

This study examined how presumptions in questions related to individuals' memory of prior events. A presumption refers to a condition in a question that must be true in order for the question to make sense. So a question like, "what color was the speeding car?" presupposes that the car was speeding. In this study, college students in small groups of 10 watched a video in which a white car causes a five-car chain reaction accident. The video was a minute long, but the car accident itself lasted only 4 seconds. After the film, the students were given a 10-item questionnaire. For half of the students, the first question asked, "How fast was the white car going?" For the other half of students, the first question contained a presumption, which asked, "How fast was the white car going when it passed the barn?" The remaining questions were irrelevant to the experiment except for the final question, which asked, "Did you see a barn in the video?" In actuality, the video did not contain a barn.

Of the group of students that were only asked about the speed of the car, only 5% claimed to have seen a barn. Of the group that was asked the presumptive question, 78% claimed to have seen a barn.

This research shows that individuals do not accurately remember previous events because presumptive questions can illicit false memories. Although simple in its design, this study demonstrates the importance of detecting presumptions when using eyewitness testimony to recount a prior event.

Children's Impulsivity and Grades (between-group study with effect size error)

This study examined how children's impulsive behaviors related to academic achievement in school. Children in 2nd grade were used for this study. The researchers tested children's impulsiveness by asking them to wait to eat an M&M that was placed on a table directly in front of them. The M&M was in a clear plastic container, which the child had to lift in order to get the treat. The experimenter, who sat on the other side of the table, told the child they had to wait until she rang a bell before they were allowed to open the container and eat the treat. If the child took the M&M before this time he/she was categorized as "impatient." Those children who waited until the experimenter rang the bell were categorized as "patient."

After controlling for a variety of variables, such as socio-economic status, child sex, language ability, and parental education, the researchers found that impatient children had significantly lower grades (C+ average) than the patient children (B-average) at the end of the school year.

These findings suggests that patient children do well in school because they are good at controlling their impulses when compared to impatient children who have little control over their impulses. Therefore, in order to improve academic achievement in elementary school, children should be taught effective self-control strategies.

Personal messages and public obedience (between-group study with effect size error)

A group of social psychologists set out to examine public obedience. They did this by observing whether or not people parked illegally in handicap spaces. After obtaining permission from the city, the experimenters observed a single parking lot with a set of four handicap spaces under two conditions: using a vertical sign or a sign that contained a personal message. The vertical sign contained the standard white handicap access symbol on a blue background. The sign containing a personal message read, "WARNING: THIS SPACE WATCHED BY CONCERNED CITIZENS." Both manipulations lasted for a week. In order to control for ordering effects, the experimenters ran the study a second time and switched the order the conditions were presented. When the drivers parked in these spaces and left, the experimenters checked to see if the vehicle had a handicap display, handicap license plates, or handicap modifications on the car. If none of these criteria were met, the car was recorded as illegally parked.

After analyzing the data, they found significant differences between the signs. People parked illegally 26% of the time when the vertical sign was displayed

whereas illegal parking occurred only 22% of the time when the personal message was displayed.

This study shows that people are more likely to obey public ordinances when they feel their actions are being watched, and more likely to disobey public ordinances when they think no one is watching them. Therefore, personalizing public signs to convey a sense of vigilant public monitoring may influence people to obey public laws and ordinances.

Anger and Task Frustration in Children (correlational study with no interpretive error)

A group of experimenters wanted to study the relationship between anger and task frustration in boys. Boys, 8 to 10 years of age, participated in the study. Anger was assessed using a 10-item scale that examined the strength to which the boys exhibited feelings of annoyance, displeasure, and or hostility in the classroom within the past week. The scale for each items ranged from a score of 1 to 5, where a score of 1 reflected no angry feelings and a 5 reflected a very angry feelings. The boys' teachers filled out the assessment. Each boy was asked to solve 4 different pencil and paper mazes. Two of the mazes were relatively easy, whereas the other two mazes were impossible to solve. To control for order effects, the mazes were ordered randomly. The experimenters assessed task frustration by recording instances where the boys would display frustrating reactions (breathing heavily or grunting loudly) when they attempted to solve the impossible mazes.

After controlling for numerous variables, such as SES, child sex, and parental education, the experimenters found a positive relationship between anger and task frustration.

This study shows that lower anger scores were associated with lower task frustration, whereas higher anger scores were associated with higher task frustration. Studies, such as these, are relevant for understanding how anger in children relates to how they deal with difficulties in the classroom.

Social engagement and memory (correlational study with no interpretive flaw)

In this study, researchers wanted to examine the links between social engagement and short-term memory in older age. A large sample of retired adults, between 70 to 80 years of age, participated in the study. In order to control for confounding influences, participants were screened for hearing and vision loss. Social engagement was assessed by asking the participants write down the names of people they routinely interacted with during a given week. The total number of people that the participant listed was used as the social engagement score. To test-short term memory, the researchers asked each participant to recall a set of digits that the researcher read aloud. This task had a total of 6 trails. The first trail consisted of 3 digits (eg., 5, 3, 9). Each subsequent trial consisted of a new set of digits with an additional digit string (e.g., 6, 9, 2, 4 for trail #2). The total number of correctly recalled numbers was used as the short-term memory score.

After controlling for education, cognitive activities, sex, and income, the study found a positive relationship between social engagement and short-term memory. Low social engagement was associated with lower short-term memory scores, whereas high social engagement was associated with higher short-term memory scores.

Although it common knowledge that mental exercise is important for maintaining good memory, this study shows that frequent social contact with others may help maintain good memory as well.

Motivation and career achievement (correlational study with correlation/causation error)

A group of researchers set out to understand how achievement motivation causes different job status outcomes among older adults. Achievement motivation is a willing desire to be successful at a skill, trade, or profession. A sample of adults in a national corporation, 40 to 48 years of age, participated in this study. All the participants came in with low-status positions and had been working for the company for the last 10 years. Participants were given the Need for Achievement Scale to fill out. This 5-point scale contains 14-question items regarding their perceived ability to produce a desired goal and their self-esteem about their achievement. Higher scores on this scale reflected higher need for personal achievement. The participants were also asked to rank the status of their current job position on a scale of 1 to 10, where 1 signified a low status position, and 10 reflected a high status position within the corporation.

After controlling for sex, level of education, and prior job experience, the experimenters found a positive relationship between achievement motivation and job status.

This study shows that people stay in low status positions because they lack the personal motivation to achieve, whereas obtaining a high status position is the end up with more successful careers when compared to people with little motivation.

The Home environment and alphabet knowledge (correlational study with correlation/causation error)

A group of researchers set out to understand how richness in the home environment influenced children's knowledge of the alphabet. To do this, the researchers visited the homes of sample of families who had a 4- to 5-year-old child. They examined the richness of the home environment by taking record of the number of children's books in the household. After this was done, children were presented with a set of flash cards, each which contained a single letter in the alphabet. Children were asked to name the letter on the flashcard. In order to control for ordering effects, the cards were shuffled prior to asking children to name each letter.

After controlling for numerous variables, such as SES, child sex, and parental education, they found a positive relationship between the number of children's books and knowledge of letters in the alphabet.

This study shows that having books at home causes children to learn more letters, whereas having fewer books at home causes children to learn fewer letters. These findings, therefore, emphasize the importance of providing young children with enriching experiences in the home.

Appendix B. Study 2 Articles

Note: The articles listed below were for the anecdote condition.

Between-group Study 1

Younger Siblings are More Creative

Jacob is only two years younger than his brother Brendan, but they couldn't be more different. While Brendan prefers to spend his free time watching TV and playing videogames, Jacob is often sketching his latest universe, creating costumes from regular household items, and speaking in imaginary languages. Their mom has often wondered if this difference is because Jacob is the younger child. Interestingly, researchers who have pondered this same question find that younger siblings are more creative.

As part of a recent study, researchers followed 542 sibling pairs. The oldest child, who was 10-12 years old, was given a standard creativity assessment. The younger sibling was assessed years later, the age when their older sibling had taken the assessment. The study found differences in creative abilities between siblings. Younger siblings' creativity scores were 2 points higher than older siblings' scores.

Based on these findings, the researchers suggest that asking about birth order may help people identify creative individuals.

Between-group Study 2

Suggestive Questions Create False Memories

Sarah witnessed a car accident on her way to work. When Sarah informed the police that she saw everything happen, the officers asked her several questions for their report. In courtrooms, eyewitness accounts are an important source of evidence. However, even though Sarah has no real reason to lie, it's hard to know whether her memories of the accident are real. That is, the things she remembered about the accident may have never happened. In an effort to understand false memories, researchers have found that questions have a powerful effect on what people *think* they remember.

In one study, researchers asked 491 college students to watch a video of a truck causing a car accident. Half of the students were asked suggestive questions, implying that there was a barn in the video (e.g., "What was the truck doing when it passed the barn?"). But in reality, no barns were present. The other half of students weren't asked suggestive questions. After being asked to remember what they saw, students who were asked suggestive questions reported seeing a barn 2 percent more than the students who weren't asked suggestive questions.

Based on these findings, the researchers recommend that officers should avoid asking suggestive questions when questioning witnesses.

Between-group Study 3

Being Overly Thankful Makes People Less Happy

Mr. McConnell starts off every dinner by asking his family to share what they are thankful for. All of them, however, never seem to benefit much from this routine. They can never think of new things to say, so their efforts are always half-hearted. Although Mr. McConnell believes doing this helps his family appreciate life, in reality, none of them feel any happier. In an effort to understand the roots of happiness, research has also found that being overly thankful yields fruitless results.

In a recent study, researchers asked 491 adults between 30-40 years of age to write about things they were thankful for (e.g., my kids, my health). Half of the participants were asked to do this five times a week, whereas the other half were asked to do this only once a week. After a month, they were asked whether they were happy with their lives. The study found that being overly thankful decreased happiness. Those who did the exercise five times a week reported feeling happy 3 percent less than those who only did the exercise once a week.

Based on these findings, the researchers warn that being too thankful can negatively affect how happy people feel.

Between-group Study 4

Having Enemies Improves Children's Social Development

Middle school students Amy and Sandra don't like each other. Amy remarks, "At first Sandra and I were friends, but she wasn't who I thought she was." Sandra notes, "I heard she was saying bad things about me. And can you believe she denied it? Unbelievable!" What's interesting about Amy and Sandra is that despite their grudge, both seem well adjusted; they have close friends and get along with their classmates. In an effort to understand how negative peer relationships impact children, research is uncovering that having enemies actually improves social development.

In this study, researchers asked 590 middle school students to privately report whether they disliked a classmate. The researchers then identified the children who disliked each other. The children with no mutual dislikes served as the comparison group. The study found that children with enemies had better social skills. For both girls and boys, children who disliked each other scored 3 points higher on a social skills assessment than children in the comparison group.

Based on these findings, the researchers advise parents to view children's negative peer relationships as opportunities for personal growth.

Correlational Study 1

Active Social Lifestyles Improves Retired Adults' Memory

Savannah and George, a lovely elderly couple, have been married for 60 years. Savannah is very popular in the retirement home and enjoys socializing with others. George, however, doesn't like people much, so he spends his time solving crosswords, reading the paper, and listening to the radio. Lately, George has been having a hard time remembering things, while Savannah's memory remains as sharp as a blade. In an effort to understand the lives of retired adults, new research is also revealing that active social lifestyles helps improve their memory.

In a recent study, researchers followed 481 retired adults between 70 and 85 years of age. They surveyed the number of social activities they engaged in during a given week and assessed their short-term memory. The study found a positive correlation between the two variables. Having an active social lifestyle increased retired adults' short-term memory. In contrast, having an inactive social lifestyle decreased their short-term memory.

Based on these findings, the researchers suggest that older adults should socialize more frequently to improve their memory.

Correlational Study 2

Controlling Mothers Increase Children's Weight Gain

Like many concerned mothers, Emily Brown wants to prevent her child, Samantha, from becoming overweight. In an effort to prevent this, however, she controls everything Samantha eats. She limits the portion of Samantha's meals and prohibits her from eating packaged snacks. Yet, Samantha has been gaining more weight than is normal. Parents like Emily feel that they are helping their child, but a new study has revealed that controlling mothers are actually increasing children's weight gain.

As part of a recent study, researchers followed 583 mothers and observed how much control they asserted over their child's eating habits. These children, who were 6-10 years old, were assessed using the body mass index scale. The study found a positive correlation between the two variables. Controlling mothers made their children gain more weight. In contrast, less controlling mothers made their children gain less weight.

Based on these findings, the researchers suggest that mothers should exert less control over their children's eating habits.

Correlational Study 3

Social Cluelessness Increases Depressive Symptoms

Jeff, who considers himself a fun and outgoing guy, suffers from depression. He tells his therapist that it's mostly because he has trouble connecting with others. Although Jeff doesn't like to admit it, he has a bad habit of upsetting or annoying people he meets. His therapist believes this happens because Jeff doesn't attend to important social cues. That is, he doesn't adjust his behaviors to match what others are expressing non-verbally. New research is also giving rise to the idea that "social cluelessness" leads to depression.

In one study, researchers asked 628 college students to fill out a social cluelessness scale, which assessed how often they failed to read a social situation (e.g., I don't realize when someone is uncomfortable until they make it clear to me). They were also given a depression scale, which assessed how often they experienced persistent sadness and loss of interest. The study found a positive correlation between the two variables. Being socially clueless increased depressive symptoms. In contrast, being socially mindful decreased depressive symptoms.

Based on these findings, the researchers suggest that being mindful in social situations may help prevent depression.

Correlational Study 4

Introverted Partners Decrease Marital Satisfaction

Amanda knew Frank was shy when she married him. Although Frank is a warm and caring person, his introverted tendencies have started to take a toll on their relationship. At social gatherings, Frank always tries to finds a quiet corner. This makes things awkward for Amanda, who doesn't know whether to sit next to him or continue to mingle with others. Amanda remarks, "Although I love him, it frustrates me when he gets like this. It makes me question our compatibility." The latest research is also revealing that introverted partners negatively impact marriages.

In a recent study, researchers followed 437 middle-aged couples who were married between 4-6 years. Both were asked to fill out an introversion scale (e.g., I am often anxious about approaching people). They also filled out a martial satisfaction scale, which assessed their satisfaction towards various dimensions of their relationship. The study found a negative correlation between both variables. For both males and females, having an introverted partner decreased marital satisfaction in the other partner. In contrast, having a partner who was more extroverted increased marital satisfaction in the other partner.

From these findings, the researchers suggest that introverted partners must learn to overcome their anxieties to maintain a healthy marriage.

Appendix C. Rational-Experiential Inventory

(Pacini & Epstein, 1999)

For each of the statements below, mark the alternative that best describes your opinion. There are no right or wrong answers so do not spend too much time deciding on an answer. The first thing that comes to mind is probably the best response. Response options: *1 – Strongly disagree, 2 – Disagree, 3 – Neither agree nor disagree, 4 – Agree, 5 – Strongly Agree*.

1. I try to avoid situations that require thinking in depth about something. (Reversed Scored)
2. I like to rely on my intuitive impressions. (Reversed Scored)
3. I'm not that good at figuring out complicated problems. (Reversed Scored)
4. I don't have a very good sense of intuition.
5. I enjoy intellectual challenges.
6. Using my gut feelings works well for me in figuring out problems in my life. (Reversed Scored)
7. I am not very good at solving problems that require careful logical analysis. (Reversed Scored)
8. I believe in trusting my hunches. (Reversed Scored)
9. I don't like to have to do a lot of thinking. (Reversed Scored)
10. Intuition can be a very useful way to solve problems. (Reversed Scored)
11. I enjoy solving problems that require hard thinking.
12. I often go by my instincts when deciding on a course of action. (Reversed Scored)
13. Thinking is not my idea of an enjoyable activity. (Reversed Scored)
14. I trust my initial feelings about people. (Reversed Scored)
15. I am not a very analytical thinker. (Reversed Scored)
16. When it comes to trusting people, I can usually rely on my gut feelings. (Reversed Scored)
17. Reasoning things out carefully is not one of my strong points. (Reversed Scored)
18. If I were to rely on my gut feelings, I would often make mistakes.
19. I prefer complex problems to simple problems.
20. I don't like situations in which I have to rely on intuition.
21. Thinking hard and for a long time about something gives me little satisfaction. (Reversed Scored)
22. I think there are times when one should rely on one's intuition. (Reversed Scored)
23. I don't reason well under pressure. (Reversed Scored)
24. I think it is foolish to make important decisions based on feelings. (Reversed Scored)
25. I am much better at figuring things out logically than most people.
26. I don't think it is a good idea to rely on one's intuition for important decisions.
27. I have a logical mind.
28. I generally don't depend on my feelings to help me make decisions.
29. I enjoy thinking in abstract terms.
30. I hardly ever go wrong when I listen to my deepest gut feelings to find an answer. (Reversed Scored)
31. I have no problem thinking things through carefully.
32. I would not want to depend on anyone who described himself or herself as intuitive.
33. Using logic usually works well for me in figuring out problems in my life.
34. My snap judgments are probably not as good as most people's.
35. Knowing the answer without having to understand the reasoning behind it is good enough for me. (Reversed Scored)
36. I tend to use my heart as a guide for my actions. (Reversed Scored)
37. I usually have clear, explainable reasons for my decisions.

38. I can usually feel when a person is right or wrong, even if I can't explain how I know. (Reversed Scored)
39. Learning new ways to think would be very appealing to me.
40. I suspect my hunches are inaccurate as often as they are accurate.

Appendix D. Actively Open-Minded Thinking Scale

(Stanovich & West, 1997)

For each of the statements below, mark the alternative that best describes your opinion. There are no right or wrong answers so do not spend too much time deciding on an answer. The first thing that comes to mind is probably the best response. Response options: *1 – Disagree strongly, 2 – Disagree moderately, 3 – Disagree slightly, 4 – Agree slightly, 5 – Agree moderately, 6 – Agree strongly*

1. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups. (Reversed Scored)
2. What beliefs you hold have more to do with your own personal character than the experiences that may have given rise to them. (Reversed Scored)
3. I tend to classify people as either for me or against me. (Reversed Scored)
4. A person should always consider new possibilities.
5. There are two kinds of people in this world: those who are for the truth and those who are against the truth. (Reversed Scored)
6. Changing your mind is a sign of weakness. (Reversed Scored)
7. I believe we should look to our religious authorities for decisions on moral issues. (Reversed Scored)
8. I think there are many wrong ways, but only one right way, to almost anything. (Reversed Scored)
9. It makes me happy and proud when someone famous holds the same beliefs that I do. (Reversed Scored)
10. Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune.
11. There are a number of people I have come to hate because of the things they stand for. (Reversed Scored)
12. Abandoning a previous belief is a sign of strong character.
13. No one can talk me out of something I know is right. (Reversed Scored)
14. Basically, I know everything I need to know about the important things in life. (Reversed Scored)
15. It is important to persevere in your beliefs even when evidence is brought to bear against them. (Reversed Scored)
16. Considering too many different opinions often leads to bad decisions. (Reversed Scored)
17. There are basically two kinds of people in this world, good and bad. (Reversed Scored)
18. I consider myself broad-minded and tolerant of other people's lifestyles.
19. Certain beliefs are just too important to abandon no matter how good a case can be made against them. (Reversed Scored)
20. Most people just don't know what's good for them. (Reversed Scored)
21. It is a noble thing when someone holds the same beliefs as their parents. (Reversed Scored)
22. Coming to decisions quickly is a sign of wisdom. (Reversed Scored)
23. I believe that loyalty to one's ideals and principles is more important than "open-mindedness." (Reversed Scored)
24. Of all the different philosophies which exist in the world there is probably only one which is correct. (Reversed Scored)
25. My beliefs would not have been very different if I had been raised by a different set of parents. (Reversed Scored)
26. If I think longer about a problem I will be more likely to solve it.
27. I believe that the different ideas of right and wrong that people in other societies have may be

valid for them.

28. Even if my environment (family, neighborhood, schools) had been different, I probably would have the same religious views. (Reversed Scored)
29. There is nothing wrong with being undecided about many issues.
30. I believe that laws and social policies should change to reflect the needs of a changing world.
31. My blood boils over whenever a person stubbornly refuses to admit he's wrong. (Reversed Scored)
32. I believe that the "new morality" of permissiveness is no morality at all. (Reversed Scored)
33. One should disregard evidence that conflicts with your established beliefs. (Reversed Scored)
34. Someone who attacks my beliefs is not insulting me personally.
35. A group which tolerates too much difference of opinion among its members cannot exist for long. (Reversed Scored)
36. Often, when people criticize me, they don't have their facts straight. (Reversed Scored)
37. Beliefs should always be revised in response to new information or evidence.
38. I think that if people don't know what they believe in by the time they're 25, there's something wrong with them. (Reversed Scored)
39. I believe letting students hear controversial speakers can only confuse and mislead them. (Reversed Scored)
40. Intuition is the best guide in making decisions. (Reversed Scored)
41. People should always take into consideration evidence that goes against their beliefs.

Appendix E. Epistemic Beliefs About Psychological Research Scale

(Estes et al., 2003)

The purpose of these following questions is to understand your attitudes towards psychological research. The following statements refer to research on *psychological* health and development. Researchers in this field try to find out how different factors influence or are related to social, emotional, and intellectual development.

1. On most issues in this field, with enough careful research, scientific experts can sooner or later be certain that their findings are correct.
    - Disagree very strongly
    - Disagree strongly
    - Disagree
    - Not sure
    - Agree
    - Agree strongly
    - Agree very strongly

2. On most issues in this filed, I would trust my own opinion more than the opinion of a scientific expert. (Reversed Scored)
    - Disagree very strongly
    - Disagree strongly
    - Disagree
    - Not sure
    - Agree
    - Agree strongly
    - Agree very strongly

3. On most issues in this field, I would trust the opinions of a respected adult over the opinion of a scientific expert. (Reversed Scored)
    - Disagree very strongly
    - Disagree strongly
    - Disagree
    - Not sure
    - Agree
    - Agree strongly
    - Agree very strongly

4. On most issues in this field, I would change my mind if I heard about new scientific evidence that went against my beliefs.
    - Disagree very strongly
    - Disagree strongly
    - Disagree
    - Not sure
    - Agree
    - Agree strongly
    - Agree very strongly

Appendix F. Study 2 Prior Beliefs Questions

Please respond to the following statements.

Response options: *1 – Disagree very strongly, 2 – Disagree strongly, 3 – disagree, 4 – No opinion, 5 – Agree, 6 – Agree strongly, 7 – Agree very strongly*

1. **"Younger Siblings are More Creative"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

2. **"Suggestive Questions Create False Memories"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

3. **"Being Overly Thankful Makes People Less Happy"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

4. **"Having Enemies Improves Children's Social Development"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

5. **"Active Social Lifestyles Improves Memories in Retired Adults"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

6. **"Controlling Mothers Increase Children's Weight Gain"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

7. **"Social Cluelessness Increases Depressive Symptoms"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

8. **"Introverted Partners Decrease Marital Satisfaction"**
   Before completing this study, this statement would've fit with my beliefs and experiences.

Appendix G. Study 2 Background Questions

1. Please list your sex.
    - Female
    - Male
2. Please list your age.
3. Please indicate your race or ethnicity (This will only be used to obtain descriptive information about our study sample).
    - African American
    - African, non-American (e.g., African, West Indian, etc.)
    - Bi-racial/Mixed/Multicultural/Multi-racial
    - Caucasian (White/European Ancestry)
    - East Asian (Chinese, Japanese, Korean, etc.)
    - Hispanic/Latino/Chicano/Puerto Rican
    - Native American
    - Pacific Islander (Filipino, Samoan, etc.)
    - South Asian (Indian, Pakistani, etc.)
    - Southeast Asian (Cambodian, Laotian, Vietnamese, etc.)
    - Other
4. Is English your primary language? (Yes, No)
5. How many years have you been a student at the University of Michigan?
    - Entering Student
    - 1 year
    - 2 years
    - 3 years
    - 4 years
    - 5 years
    - 6 years
    - More than 7 years
6. What is your current class standing?
    - Freshmen
    - Sophomore
    - Junior
    - Senior

7. Have you taken or are currently taking a basic statistics or research methods course at Michigan (100-200 level courses)? (Yes, No)

8. Have you taken or are currently taking an upper-division statistics or research methods course at Michigan (300-400 level courses)? (Yes, No)

9. Please report your composite SAT score. If you don't fully remember, please estimate what you think you received. If you didn't take the SAT, mark "N/A."

10. Please report your composite ACT score. If you don't fully remember, please estimate what you think you received. If you didn't take the ACT, mark "N/A."

References

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., et al. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. *Cognitive Development, 23,* 452-471.

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakable beliefs. *Journal of Experimental Social Psychology, 19*, 93-108.

APA. (2009). *Publication Manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Arum, R., & Roska, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago: The University of Chicago Press.

Baesler, E. J., & Burgonn, J. K. (1994). The temporal effects of story and statistical evidence on belief change. *Communication Research, 21*(5), 582-602.

Baron, J. (2008). *Thinking and deciding* (4th ed.). Cambridge: Cambridge University Press.

Bastardi, A., Uhlmann, E. L., & Ross, L. (in press). Wishful thinking: belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*.

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.

Bekker, H. L., Hewison, J., & Thorton, J. G. (2003). Understanding why decision aids work: Linking process and outcome. *Patient Education & Counselling, 50*, 323-329.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing*. New York: Basic Books.

Bensley, A. D. (2010). A brief guide for teaching and assessing critical thinking in psychology. *Observer*, *23*(10), 49-53.

Bensley, A. D., Crowe, D. S., Bernhardt, P., Buckner, C., & Allman, A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in psychology. *Teaching of Psychology, 37*(2), 91-96.

Bensley, A. D., & Haynes, C. (1995). The acquisition of general purpose strategic knowledge for argumentation. *Teaching of Psychology, 22*(1), 41-45.

Beyerstein, B. L. (2001). Alternative medicine and common errors of reasoning. *Academic Medicine, 76*, 230-237.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*, 44-48.

Bullock, M., Sodian, B., & Koerber, S. (2008). Doing experiments and understanding science. Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from a 20 year longitudinal study*. Mahwah, NJ: Erlbaum.

Burrage, M. S. (2008). *"That's an interesting finding, but....:" Postsecondary students' interpretations of research findings.* Unpublished Doctoral Dissertation, The University of Michigan, Ann Arbor, MI.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45*(4), 805-818.

Chi, M. T. H., Glaser, R., & Reese, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.

Churchman, C. W. (1971). *The design of inquiring systems: Basic concepts of systems and organizations*. New York: Basic Books.

Dahlstrom, M. F. (2010). the role of causality in information acceptance in narratives: An example from science communication. *Communication Research, 37*(6), 857-875.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*(5), 428-433.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*(5), 483-489.

Dressel, P. L., & Mayhew, L. B. (1954). *General education: Explorations in evaluation*. Washington, D.C.: American Council on Education.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267-285.

Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review, 32*(1), 81-111.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4-10.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconcious. *American Psychologist, 49*(8), 709-724.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*(2), 390-405.

Estes, D., Chandler, M., Horvath, K. J., & Backus, D. W. (2003). American and British college students' epistemological beliefs about research on psychological and biological development. *Applied Developmental Psychology, 23*, 625-642.

Evans, J. S. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning, 4*(1), 45-110.

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science*, *7*(10), 454-469.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning, 13*(4), 321-339.

Evans, J. S. B. T., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. New York: Oxford University Press.

Evans, J. S. B. T., Handley, S. J., & Harper, C. N. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology, A: Human Experimental Psychology, 54*(3), 935-958.

Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decsions: Is a picture worth a thousand words? *Medical Decision Making, 25*, 398-405.

Feldman, K. A., & Newcomb, T. M. (1969). *The impact of college on students*. San Francisco: Jossey-Bass.

Foley, R. (1987). *A theory of epistemic rationality*. Cambridge, MA: Harvard University Press.

Follmer, G. A., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology, 18*, 203-221.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253-292.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34-45.

Giancarlo, C. A., & Facione, P. A. (2001). A look across four years at the disposition towards critical thinking among undergraduate students. *The Journal of General Education, 50*(1), 29-55.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650-669.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Gobet, F., & Simon, H. (1996). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory and Cognition, 24*, 493-503.

Gobet, F., & Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving from grand-master-level chess. *Psychological Science, 7*(1), 52-55.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75-90.

Graesser, A. C., Singer, M., & Trabasco, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395.

Halpern, D. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449-455.

Halpern, D. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning, 80*, 69-74.

Hatfield, J., Faunce, G. J., & Job, S. (2006). Avoiding confusion surrounding the phrase "correlation does not imply causation". *Teaching of Psychology, 33*(1), 49-51.

Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation, 15*, 425-437.

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist, 39*(1), 43-55.

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.

Hofer, B. K., & Pintrich, P. R. (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515-526.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251.

Karabenick, S. A., & Moosa, S. (2007). Culture and personal epistemology: U.S. and Middle Eastern students' beliefs about scientific knowledge and knowing. *Social Psychology of Education, 8*(4), 375-393.

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology, 88*(2), 260-271.

King, P. M. (2000). Learning to make reflective judgments. *New Directions for Teaching and Learning, 82*, 15-26.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass Publishers.

King, P. M., Wood, P. K., & Mines, R. A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13*(2), 167-186.

Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology*, *2*, 89-116.

Klaczynski, P. A. (2001a). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child Development, 72*(3), 844-861.

Klaczynski, P. A. (2001b). Framing effects on adolescent task representations, analytic and heuristic processing, and decision making: Implications for the normative/descriptive gap. *Applied Developmental Psychology*, *22*, 289-309.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology*, *89*(3), 470-485.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging*, *14*(3), 400-416.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge: MIT Press.

Kosonen, P., & Winne, P. H. (1995). Effects of teaching statistical laws on reasoning about everyday problems. *Journal of Educational Psychology, 87*(1), 33-46.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155-178.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2), 16-26, 46.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review, 4*, 1-6.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23*, 435-451.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. (1980a). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. (1980b). Models of competence in solving physics problems. *Cognitive Science, 4*, 317-345.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43*(6), 431-442.

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology, 26*(6), 952-960.

Lehmann, I. J. (1963). Changes in critical thinking, attitudes, and values from freshman to senior years. *Journal of Educational Psychology, 54*(6), 305-315.

Leshowitz, B. (1989). It is time we did something about scientific illiteracy. *American Psychologist, 44*(8), 1159-1160.

Leshowitz, B., DiCerbo, K. E., & Okun, M. (2002). Effects of instruction in methodological reasoning on information evaluation. *Teaching of Psychology, 29*(1), 5-9.

Leshowitz, B., & Okun, M. (2011). Effects of an evidence-based text on scepticism, methodological reasoning, values, and juror decision-making. *Educational Psychology, 31*(3), 319-337.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences, 17*, 115-127.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107*, 343-352.

McPeck, J. E. (1981). *Critical thinking and education*. Oxford: Martin Robertson.

Mill, D., Gray, T., & Mandel, D. R. (1994). Influence of research methods and statistics courses on everyday reasoning, critical abilities, and belief in unsubstantiated phenomena. *Canadian Journal of Behavioural Science, 26*(2), 246-258.

Muis, K. R., Benedixon, L. D., & Haerle, F. C. (2006). Domain-Generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review, 18*(1), 3-54.

Nair, K. U., & Ramnarayan, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality, 34*(3), 305-328.

Nelson Laird, T. F. (2005). College students' experiences with diversity and their effects on academic self-confidence, social agency, and disposition toward critical thinking. *Research in Higher Education, 46*(4), 365-387.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238,* 625-631.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Norcross, J. C., Gerrity, D. M., & Hogan, E. M. (1993). Some outcomes and lessons from a cross-sectional evaluation of psychology undergraduates. *Teaching of Psychology*, *20*(2), 93-96.

Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher, 18*(9), 21-26.

Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30*(3), 286-313.

Nussbaum, E. M., & Bendixen, L. D. (2003). Approaching and avoiding arguments: The role of epistemological beliefs, need for cognition, and extraverted personality traits. *Contemporary Educational Psychology, 28*(4), 573-595.

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157-169.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Personality and Social Psychology Bulletin, 76*(6), 972-987.

Pascarella, E. T., Bohr, L., Nora, A., & Terenzini, P. T. (1996). Is differential exposure to college linked to the development of critical thinking? *Research in Higher Education, 37*(2), 159-174.

Perry, W. G. (1970). *Forms of intellectual and ethical development*. New York: Henry Hold and Company, Inc.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.

Pierro, A., Mannetti, L., Erb, H.-P., Spiegel, S., & Kruglanski, A. (2005). Informaion length and order of presentation as determinants of persuasion. *Journal of Experimental Social Psychology, 41*, 458-469.

Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 389-414). Mahwah, NJ: Lawrence Erlbam Associates.

Pollard, P., Newstead, S. E., Evans, J. S. B. T., & Allen, J. L. (1994). Debiasing by instruction - the case of belief bias. *European Journal of Cognitive Psychology, 6*(3), 263-285.

Rickert, R. K. (1967). Developing critical thinking. *Science Education*, *51*(1), 24-27.

Sá, W. C., Kelley, C. N., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences*, *38*, 1149-1161.

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*(3), 497-510.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89*(4), 634-656.

Sandoval, W. A. (2009). In defense of clarity in the study of personal epistemology. *The Journal of the Learning Sciences, 18*, 150-161.

Sandoval, W. A., & Milwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*, 498-504.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology, 85*(3), 406-411.

Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded model and coordinated research approach. *Educational Psychologist, 39*(1), 19-29.

Schraw, G., Benedixon, L. D., & Dunkle, M. E. (2002). Development and validation of the epistemic beliefs inventory (EBI). In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schraw, G., Dunkle, M. E., & Benedixon, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology, 9*, 523-538.

Schunn, C., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23*(3), 337-370.

See, Y. H., Petty, R. E., & Evans, L. M. (2009). The impact of percieved message complexity and need for cognition on information processing and attitudes. *Journal of Research in Personality, 43*, 880-889.

Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education, 36*(2), 207-227.

Sodian, B., & Bullock, M. (2008). Scientific reasoning - Where are we now? *Cognitive Development, 23*, 431-434.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven, CT: Yale University Press.

Stanovich, K. E. (2010). *Decision making and rationality in the modern world*. New York: Oxford University Press.

Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D. Priess & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development*. New York: Springer.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2), 342-357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences, 23,* 645-726.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking and Reasoning, 13*(3), 225-247.
Stanovich, K. E., & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking and Reasoning, 14*(2), 129-167.
Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. Sternberg (Ed.), *Handbook of Human Intelligence* (pp. 225-307). New York: Cambridge University Press.
Sternberg, R. J. (Ed.). (2002). *Why smart people can be so stupid*. New Haven, CT: Yale University Press.
Stewart, S., & Dempsey, L. F. (2005). A longitudinal study of baccalaureate nursing students' critical thinking dispositions. *Journal of Nursing Education, 44*(2), 81-84.
Strange, J. J., & Leung, C. C. (1999). How anecdotal accounts in news and in fiction can influence judgments of a social problem's urgency, causes, and cures. *Personality and Social Psychology Bulletin, 25*(4), 436-449.
Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in Higher Education, 36*(1), 23-39.
Thompson, V., Evans, J. S. B. T., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language, 53*, 238-257.
Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generalizability of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 94*(1), 197-209.
Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology, 17*, 851-860.
Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Hueristics and biases. *Science, 185*, 1124-1131.
Watson, G., & Glaser, E. M. (1964). *Watson-Glazer Critical Thinking Appraisal Manual*. New York: Harcourt Brace & Company.
Weinstock, M. (2009). Relative expertise in an everyday reasoning task: Epistemic understanding, problem representation, and reasoning competence. *Learning and Individual Differences, 19*, 423-434.
Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*(3), 470-477.
West, R. F., Toplak, M., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930-941.
Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review, 11*(4), 411-427.
Winterbottom, A., Bekker, H., Conner, M., & Mooney, A. (2008). Does narrative information bias individual's decision making? A systematic review. *Social Science & Medicine, 57*, 2079-2088.
Zeineddin, A., & Abd-El-Khalick. (2008). On coordinating theory with evidence: The role of epistemic commitments in scientific reasoning among college students.

Eurasia Journal of Mathematics, Science & Technology Education, 4(2), 153-168.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99-149.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172-223.