“That’s an interesting finding, but….:”
Postsecondary students’ interpretations of research findings

by

Marie S. Burrage

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education and Psychology)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Priti R. Shah, Chair
Professor Daniel P. Keating
Professor Kevin F. Miller
Research Scientist Stuart A. Karabenick

Acknowledgements

I'd like to thank my dissertation committee for their feedback and advice as I work on this project. I'd especially like to thank the chair of my committee and my advisor, Priti Shah. Without Priti's help and support, this dissertation would not have been possible. Thank you to my family, friends and Dan for their love and support throughout this process.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Two studies examined individuals' open-ended interpretations of research findings in the context of Deanna Kuhn's model of theory-evidence coordination. The first study compared undergraduate and graduate students' explanations and evaluations of research findings in familiar and unfamiliar domains. The second study considered the role of epistemic beliefs and thinking dispositions on less and more advanced college students' interpretations of research findings.

Results from both studies indicated that some people tended to focus more on the evaluation of evidence presented to them while others accepted the evidence at face value and focused on explanation, attempting to "make sense" of the research findings in terms of what they already knew (prior theories and beliefs). Educational experience was related to people's interpretation of the research findings. More advanced students tended to describe and critically evaluate data while less advanced students explained the data without considering their quality. Additionally, there were differences in the way students with different majors/areas of study interpreted research findings, with history and engineering students providing more evaluations of the findings than psychology students. Finally, correlational and regression analyses indicated that statistical/methodological training, general intelligence and sophisticated thinking dispositions were correlated with people's tendency to evaluate research findings.

Chapter 1

Introduction

Importance of Evidence Interpretation Skills

In today's "Information Age" individuals are bombarded with data that they must use to make reasoned decisions in a broad range of contexts, such as what kind of medical treatment to select for oneself or loved one, who to vote for, and what brand of child safety seat to purchase. Nonetheless, even highly educated adults have difficulty reasoning with data. Halpern (1998) reviewed literature indicating that large percentages (70-99%) of people in the American public believe in paranormal phenomena for which the evidence is sparse and flawed. For example, in a study of college students, more than 99% claimed to believe in at least one of many paranormal phenomena including channeling, psychic healing, and UFOs, with more than 65% reporting that they personally experienced at least one of these things (Messer & Griggs, 1989).

Even expert scientists make reasoning errors. Studies of trained psychologists have shown that, while the psychologists are quick to criticize experimental designs with small samples sizes, they will accept their own anecdotal and personal experiences as credible sources of data (Dawes, 1994). Many people do not seek out data that might disprove their strongly held beliefs; however, even those who do are likely to struggle with interpretation of these data. For example, a *Life* magazine article (Miller, 1997) described the experience of a man described as a physicist-astrologer. This man claimed that he considered astrology to be "useless poppycock" (p. 46) until he began to see the data; however, the data that changed his mind were “a few, small, but significant correlations” (p. 46) amongst many non-significant correlations. In this case, it was not a lack of data that led this man to his faulty conclusions, it was a misinterpretation of the data he did have.

Although careful interpretation of evidence is a crucial aspect of everyday critical thinking and scientific reasoning, people clearly struggle with this process. Therefore, a fundamental goal of both K-12 and higher education is improving these skills in students (Halpern, 2001; Solomon & Perkins, 1989). To this end, improving everyday scientific reasoning skills is a stated goal of the "National Educational Goals" promoted by the National Governor's Council and President Bush (Halpern, 2001).

There is a wide range of errors that people make when interpreting evidence. In the current project, the evidence presented to people takes the form of research findings. I decided to focus on this domain because it is relevant to everyday, real world reasoning contexts. People are consistently exposed to new research findings in popular media outlets such as newspaper articles and the nightly news, and they must interpret these research findings to the best of their abilities. As in with any new piece of information, people interpret new research finding in light of what they already know and believe (Vosniadou, 1994). Most of the research findings presented in popular media outlets concern topics about which people have preexisting beliefs, and one common issue is that, when interpreting research findings, people tend to let their prior beliefs and theories interfere with their ability to evaluate the data at hand (Klaczynski, 2000). Because of this, they may not recognize a wide range of potential problems with the evidence, ranging from flaws in the experimental methodology to over-interpretation of the research findings by the author of the piece.

For example, when interpreting experimental data, people in Klaczynski's (2000) study fell into a belief bias trap that left them "blind" to explicit validity threats, selection bias, and a questionable construct. For example, when upper middle class participants were presented with a description of an experiment indicating that the upper middle class was somehow "superior" in some way to the working class, they tended not to recognize that upper middle class subjects had been selected in a very different context (they were in a business setting) than the working class subjects (they were at a rally).

Other research has focused on the difficulties that people have with specific kinds of statistical and methodological reasoning problems. For example, Hatfield and Faunce (2006) describe the confusion that often arises when interpreting correlational data. They emphasize that students have problems learning to distinguish between the concepts of correlation and causation in the context of statistical and methodological courses. Additionally, Barry Leshowitz (1989) examined the methodological reasoning of undergraduate students, and found that, before participating in a psychology research methods class, the students had trouble with topics such as operational definitions, sampling, predictive relationships and controlled comparisons. However, to my knowledge, there isn't a body of research that compares people's recognition of these different types of statistical and methodological errors.

In the current project, I propose to examine the processes that people engage in when interpreting research findings using the model described below. Because people are frequently exposed to research findings and scientific data in their everyday lives, it is crucial that I gain a better understanding of how they are interpreting this information, especially when the information is flawed in some way. Therefore, it is a primary goal of this project to apply Deanna Kuhn's research on evidence vs. explanation-based reasoning in this new domain with the hopes of developing a more systematic way of thinking about people's interpretation of research findings. A second goal of the current research is to investigate individual factors, such as intelligence, thinking dispositions and amount and type of higher education (especially in the domains of statistics and research methodology), that contribute to stronger and weaker evidence interpretation processes. Hopefully by gaining insight into what people do when presented with research findings and why, we can move towards a better understanding of how evidence evaluation processes might ultimately be improved.

Theory-Evidence Coordination Model

Research on justifying knowledge claims has shown that people tend to focus on providing theoretically-based explanations of why their claims are valid as opposed to providing evidence that supports their claims. In other words, people tend to prefer to answer the question "Why is it so?" as opposed to "How do you know?" when justifying a claim. Deanna Kuhn and others have also found that, in certain circumstances, these explanations can lead to overconfidence, can inhibit examination of alternatives and can be false (Kuhn, 2001).

For example, in her research on jury decision-making, Kuhn and her colleagues have explored the way that jurors make and justify claims about verdict choices. They have found that jurors commonly rely on narrative explanations of what happened and select their verdict choices based on that explanation; however, they did find individual variation in use of evidence to justify claims. In other research, Kuhn and others have examined individual difference factors that they believe contribute to the tendency to engage in evidence-based reasoning. In a study with both young children and adult jurors, she has found that those with more sophisticated epistemic beliefs, tended to rely more heavily on evidence that those with less sophisticated epistemic beliefs (Kuhn, Weinstock, & Flaton, 1994). The role of epistemic beliefs in explanation vs. evidenced-based reasoning is discussed in greater detail in the following section of this introduction. Kuhn and others have also found that those with greater amounts of higher education (e.g. advanced college students) are more likely to prefer evidence to explanation (Brem & Rips, 2000; Kuhn, 2001). Kuhn's research on evidence-based reasoning is framed in terms of a model of theory-evidence coordination.

Kuhn's model of *theory-evidence coordination*, which she claims is a cornerstone of scientific reasoning as well as critical thinking more generally, outlines one way that people can reason using theory and evidence. For those using a theory-evidence coordination model to select/reconsider a theory, evidence figures heavily, multiple alternatives are considered, and the alternative that has the highest quality and most consistent evidence associated with it is the alternative that is chosen. This stands in contrast to a *satisficing* model, in which the construction of a plausible narrative is sufficient to dictate a corresponding choice, evidence inconsistent with this narrative is disregarded, and alternatives are not considered. These two models dictate very different roles for evidence in reasoning. According to Kuhn, in order to reason competently, whether it be in the domain of legal reasoning, scientific reasoning, or informal reasoning, one must be able to coordinate one's theories with new evidence bearing on them. In order to do this, one must have the ability to reflect on his/her theories as objects of cognition (distinct from external sources of evidence) to an extent sufficient to recognize that they could be wrong. Also, one must be able to recognize external sources of evidence that could disconfirm his/her theory and to carefully evaluate that evidence. Therefore, one must understand and value the role of quality evidence in proving or disproving a theory.

Many reasoning errors arise from *belief bias*, defined as individuals' inability to separate their prior expectations and beliefs from the quality of information presented to them. These errors are closely related to the tendency to reason in an overly theory-driven way. For example, belief-bias occurs when people are more critical of evidence that they do not believe or expect, and are less critical of evidence that they do believe (Edwards & Smith, 1996; Klaczynski, 2000; Lord, Ross & Leper, 1979). They are more likely to evaluate arguments as demonstrating solid reasoning when they agree with them than when they do not. And, they find it easier to generate arguments supporting their viewpoints than to generate counter-arguments (Toplak & Stanovich, 2003). Research on belief-bias has found that those operating in more advanced stages of epistemic development (i.e. are able to effectively distinguish between subjective and objective components of knowing) are less likely to fall into belief-bias traps (Stanovich & West, 1997).

Epistemic beliefs and thinking dispositions

Virtually all critical thinking researchers would agree that one's thoughts and beliefs about knowledge are predictors of one's ability to think critically. Deanna Kuhn's work has gone so far as to call thinking "an epistemological enterprise." Her work emphasizes the importance of reflection on one's own thinking process to producing good thinking, and she draws from work on the development of epistemological understanding by Perry and King and Kitchner. Kuhn's research shows that there is a clear relationship between epistemological beliefs and quality of argument skills. Additionally, Kuhn reviews research demonstrating that epistemological beliefs influence the ways that people both use their intellectual skills, and acquire new knowledge.

Kuhn presents a stage model of epistemological beliefs with four basic levels: realist, absolutist, multiplist and evaluative. The original version of this theory was primarily concerned with the development of epistemological beliefs over the college years, centered on development of beliefs about of how knowledge and truth are constructed. When one is at the realist level, one believes that knowledge is certain, easily knowable and comes from a single expert source; therefore, evidence evaluation is unnecessary. At the absolutist level, evidence evaluation becomes a tool for "comparing assertions to reality and determining their truth or falsehood." At this level, one believes that reality is still directly knowable. However, as one progresses to the multiplist level, the certainty of knowledge and reality is called into question. One comes to believe that knowledge is a product of human minds, and is, therefore, subjective and uncertain. At this stage evidence evaluation becomes irrelevant because there is no way of comparing assertions. Finally, at the evaluative level, critical thinking and evidence evaluation are once again valued as a tool that allows one to come up with sound assertions, examine multiple perspectives, and find informed answers to one's questions. Only at the most advanced, evaluative level is knowledge seen to consist of claims, which require support in a framework of alternatives, evidence, and argument (Chandler, Boyes, & Ball, 1990;

Kuhn, 1999). The cognitive task underlying this evolution is the coordination of the objective and subjective components of knowing (Kuhn, 2001). This is what allows one to use an appropriate mix of theory and evidence to justify a claim.

Kuhn makes the argument that people's epistemic beliefs shape intellectual values and hence the disposition to utilize intellectual skills (Kuhn, 2001). Therefore, thinking dispositions act to link epistemic beliefs and abilities in the context of critical thinking. While critical thinking researchers do not deny the role of intelligence in producing critical thinking, they argue that intelligence needs be accompanied by dispositions to use that intelligence to engage in critical thinking (Kuhn, 2001; Stanovich, 1997). It is the combination of general intelligence and thinking dispositions that produce the best critical thinking.

Baron (1987) introduces a theory of rationality and intelligence that emphasizes the role of thinking dispositions in producing good thinking and reasoning. In Baron's view, thinking dispositions are more malleable than cognitive abilities, and are, therefore, more teachable. He thinks of them as cognitive styles, although they are also abilities in that they reflect a definition of how to be successful. Baron, Badgio and Gaskins (1986) show that one consistent characteristic of poor students is that they tend to be defensive of their incorrect beliefs, whereas good students are able to remain open to alternative views and criticism. Actively open-minded thinking is central to Baron's concept of critical thinking, and, therefore, dispositions to be sensitive and responsive to new information, and search out and examine multiple points of view are crucial to his understanding of what allows an individual to engage in rational thinking.

In their work on reasoning biases, Stanovich (e.g. Stanovich& West, 1997) and Klaczynski (e.g. Klaczynski, 2000) have explored the role of thinking dispositions in avoiding some of the most common critical thinking errors. Their research has shown that high scores on actively open-minded thinking as well as on need for cognition (Caccioppo & Petty, 1982) positively predict ability to successfully avoid belief-biased reasoning, an important component of critical thinking. Stanovich's concept of actively open-minded thinking stands in contrast to absolutism (belief in one right answer), dogmatism, and categorical thinking. These are similar to belief concepts identified in Kuhn's model of epistemological belief development. Need for cognition is a strong motivation to engage in intellectually challenging thought and activity (Caccioppo & Petty, 1982), and it also thought to be important to one's ability to think critically (Baron, 1987).

The current project attempts to apply research on epistemic beliefs and thinking dispositions to the domain of interpretation of research findings. It follows from the evidence presented above, indicating that sophisticated epistemic beliefs and thinking dispositions such as actively open-minded thinking and need for cognition are predictive of people's ability to think critically in a wide variety of contexts, that these belief and dispositions will be predictive of people's ability to think critically about research findings. According to Kuhn's model of theory-evidence coordination, thinking critically about research findings would involve paying attention to the evidence at hand and evaluating it carefully. In this research, I attempt to determine whether evaluative epistemic beliefs and actively open-minded, cognitively engaged thinking dispositions are predictive of people's evidence-based reasoning about research findings.

Higher education

Research has shown that amount of higher education is related to sophisticated epistemic beliefs and to the ability to think critically. For example, Jehng, Johnson & Anderson (1993) conducted a study looking at students' epistemological beliefs as a function of their education level and field of study. They found that students in the "soft" fields such as social sciences, arts and humanities had more sophisticated epistemic beliefs than those in the "hard" sciences, tending to believe that knowledge is uncertain and to rely on their independent reasoning skills. They also found that graduate students possessed more sophisticated epistemic beliefs than did undergraduates. In another example, Toplak and Stanovich (2003) conducted a study examining belief-bias in undergraduate students. They found that the amount of belief bias that the students demonstrated decreased systematically with year in university. Finally, Brem and Rips found that the preference for explanations (over evidence) disappears in highly able college students under certain contexts (Brem& Rips, 2000).

Research has also shown that discipline-specific higher educational experiences differentially affect statistical and methodological reasoning. Lehman, Lempert & Nisbett (1988) and Lehman & Nisbett (1990) conducted two studies showing that both undergraduate and graduate education have significant effects on reasoning ability. In their 1988 study, they found that graduate training in both psychology and medicine significantly positively affected statistical and methodological reasoning, and psychology, medical and law training significantly positively affected conditional reasoning, while chemistry training had no effect on any of the types of reasoning studied. In their 1990 study, they found that undergraduate social science training significantly positively affected statistical and methodological reasoning, whereas natural science and humanities training produced smaller, but still marginally significant, effects. Natural science and humanities training significantly positively affected ability to reason about problems in conditional logic, whereas social science training did not. In the current study I are particularly interested in statistical and methodological training and people's ability to recognize statistical and methodological errors that I build into my task in study 2.

Current Research

The current research extends work on explanation vs. evidence-based reasoning to an area that is relevant to the field of critical thinking as well as scientific reasoning: interpretation of research findings. The set of skills that I am interested in is required of people on a day-to-day basis when reading the newspaper or watching the news. People encounter findings from research studies in many popular media outlets, and they must interpret these results based on the small amount of information provided in the article or news story. I am interested in better understanding how and why people proceed with their interpretations of research findings, and I use Kuhn's model of theory-evidence coordination to frame my research questions and hypotheses. Kuhn's model emphasizes that, in order to think scientifically, one must be able to distinguish theory from evidence and must be able to examine evidence independent of prior belief and theories. If one either disregards or readily accepts evidence, one misses a crucial step in the reasoning process (Kuhn, 2001).

Previous research has primarily focused on people's attempts to justify claims by creating or identifying explanations or evidence (Kuhn, 2001). But what happens when the evidence already exists as in the case of interpretation of experimental results? Are people able to carefully evaluate the evidence presented or do they disregard or accept the evidence and jump right to the level of explanation (regardless of the quality of the evidence)? Klaczynski's work on experiment evaluation has focused on belief-biased reasoning and cognitive development, and not on evidence vs. explanation-based reasoning per se. Therefore, the theory-evidence coordination model is ripe for application in this context. The issue of evidence vs. explanation-based reasoning is of central importance when considering how people interpret research findings on a day-to-day basis. If people do not recognize the importance of evaluating research findings before jumping to level of theory-based explanations, they are in danger of engaging in the purely theory-based reasoning that characterizes Kuhn's satisficing model. Those who are able first to evaluate research findings independently of their own theories and beliefs are truly engaging in the careful, critical reasoning that characterizes Kuhn's theory-evidence coordination model.

Sufficient evaluation of evidence is a crucial component of the theory-evidence coordination model because this model characterizes reasoners who select alternatives based on the consistency and quality of evidence that support them as opposed to alternatives that fit with their own prior beliefs and theories (and may or may not be supported by evidence) (Kuhn, 2001). These reasoners must be able evaluate the quality of different forms evidence as part of their broader reasoning and decision-making process.

In additional to being primarily focused on creation of evidence or explanations in order to justify claims, Kuhn's previous work on evidence vs. explanation-based reasoning has been heavily focused on the domain of legal reasoning (Kuhn, 2001). Two studies here build on the research on evidence vs. explanation-based reasoning and apply the model of theory-evidence coordination as lens through which to examine the ways in which people interpret research findings in everyday contexts. As discussed above, people are exposed to research findings on a regular basis in popular media, and the way that they make sense of these research findings is not well understood. Therefore, the current project is designed to improve our understanding of how people interpret research findings and why using Kuhn's model of theory-evidence coordination. The studies described in this paper address the following broad questions:

1. To what extent are Kuhn's findings on evidence-based vs. explanation-based reasoning applicable to people's interpretation of research findings?
2. To what extent is higher educational training, especially training in statistical and methodological reasoning, associated with sophisticated reasoning about research findings?
3. To what extent are epistemic beliefs and thinking dispositions involved in scientific reasoning about research findings?

The results of my two studies suggest that some people tend to focus more on the evaluation of evidence presented to them while others accept the evidence at face value and jump right to the level of explanation, attempting to make sense of the research findings in terms of what they already know (prior theories and beliefs), even when flaws are deliberately built into the evidence. My results also indicate positive relationships between evidence-based reasoning and higher educational experience, as well as epistemic beliefs, thinking dispositions and general intelligence.

Chapter 2

Study 1

Introduction

The goal of study 1 is to apply Kuhn's model of theory-evidence coordination to students' interpretation of graphs and vignettes describing research findings. Study 1 also addresses the relationship between higher education training and evidence-based reasoning. Previous studies have examined the effects of higher educational training on various aspects of critical thinking and scientific reasoning such as belief-bias avoidance, development of epistemic beliefs and statistical and methodological reasoning (e.g. Brem& Rips, 2000). Additionally, numerous researchers have studied the development of epistemic beliefs in college and graduate students (Jehng et al., 1993; King & Kitchener, 2002). In their cross-sectional and longitudinal studies with studies with graduate students (1988), Lehman and Nisbett found that graduate training in psychology and medicine positively affected statistical and methodological reasoning skills.

However, to my knowledge, no one has studied college and graduate students' open-ended interpretations of research findings using a model of evidence vs. explanation-based reasoning. In my task, participants are not given much guidance about how to respond to the research findings, and therefore I capture their first responses to the vignettes describing experimental results. These responses can take various forms, and, therefore, the data are potentially rich and informative, but also challenging to interpret.

Using Kuhn's research on evidence vs. explanation-based reasoning, I am able to make sense of this rich, authentic data. This task and method is unique in that it allows me to study the *process* by which students' reason about the kinds of data they encounter everyday, not merely a targeted output.

In order to do this, I created a coding scheme informed by Kuhn's evidence vs. explanation-based reasoning model that I used to code Introductory Psychology students' as well as graduate students' open-ended responses to graphically presented research findings. This coding scheme allowed me to determine whether students' responses were evidence or explanation-based (i.e. whether they were focused on the data presented to them or whether they accepted the data at face value and jumped right to the level of explanation). Those identified as engaging in evidence-based reasoning provided interpretations focused on the research findings themselves, providing comments on or criticism of the way the research was conducted, the nature of the data collected and/or the conclusions drawn by the researchers. Those identified as engaging in explanation-based reasoning provided interpretations focused on relating the research findings to their own theories, beliefs and/or experiences.

By using cross-sectional methods similar to those used in the Lehman, Lempert & Nisbett (1988), I tested early-stage undergraduates and graduate students from three different disciplines, with different amounts and types of scientific training. This design allows me to address the question of whether higher education is associated with a shift from explanation/theory-based to evidence-based reasoning. I selected graduate students from disciplines I believed provide different types of training that might affect their ability to reason about research findings.

Firstly, Lehman, Lempert & Nisbett's 1988 study with graduate students showed that psychology graduate students have more sophisticated statistical and methodological reasoning skills. Additionally, the vignettes/graphs that I used in my tasks described different types of behavior research using methodologies that would be particularly familiar to trained psychologists. Therefore, I recruited two groups of psychology graduate students (social psychology and cognitive psychology) to participate in this study. I predicted that, because of their knowledge of statistics and research methodology, they would provide the most sophisticated descriptions of the patterns of results when compared with other participants.

Deanna Kuhn has conducted research showing that historians coordinate theory and evidence when they engage in historical reasoning (Kuhn, Weinstock & Flaton, 1994a). They are trained to look for specific points of evidence to support their theories, and, in this way, they are similar to other social scientists. A group of history graduate students was recruited for this study, and, informed by this research, I predicted that the history graduate student's would engage in evidence-based reasoning. However, due to their lack of statistical and methodological training as compared to psychology graduate students, I predicted that the history graduate students would provide descriptions or evaluations of the data not as well-informed by statistical and methodological knowledge as those of psychology graduate students.

An additional comparison group of graduate students was recruited (mechanical engineers) as well as a large group of mostly early-stage undergraduate students. Previous research has shown that advance higher educational training is associated with sophisticated epistemic beliefs (Jehng et al., 1993; King & Kitchener, 2002) as well as strong critical thinking (Toplak & Stanovich, 2003) Therefore, I predicted that all four groups of graduate students would engage in more evidence-based reasoning that undergraduates in this study based on the fact that they had accumulated more educational training.

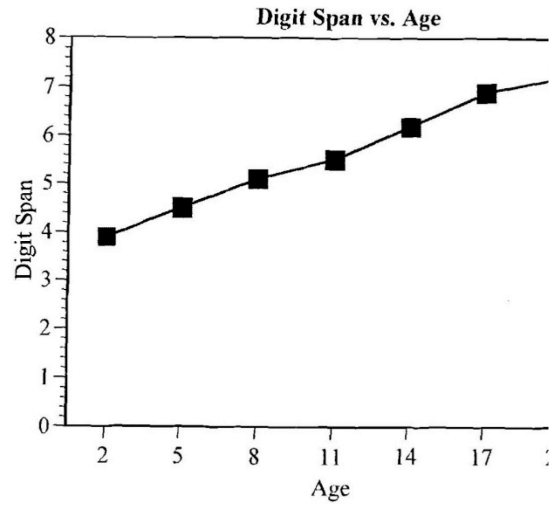
I also manipulated the content of the vignettes/graphs to test for effects of content familiarity on people's evidence vs. explanation-based reasoning. Half of the vignettes used in this study focused on general knowledge content (e.g. car accidents), whereas the other half focused on content primarily familiar to cognitive psychology graduate students (e.g. reaction time). This way, I was able to determine whether familiarity of content had an effect on people's reasoning about research findings.

Methodology

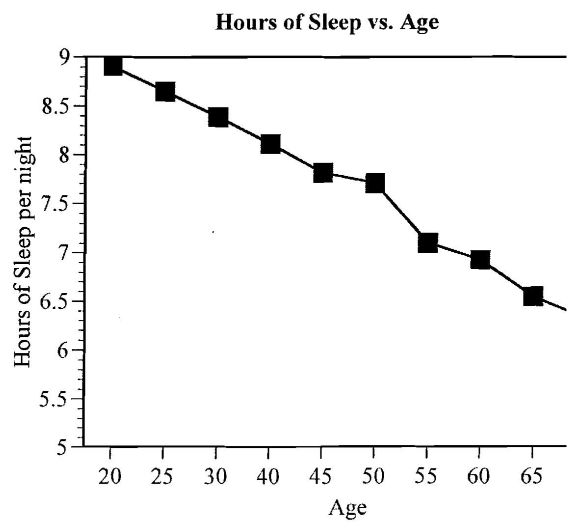
Forty-eight undergraduate students from the University of Michigan were recruited through their Introductory Psychology course. They received course credit or payment of $15 for their participation. Graduate students in cognitive psychology (n=9), social psychology (n=6), history (n=8) and mechanical engineering (n=10) were recruited via an e-mail sent to every graduate student their departments. They received $10 in exchange for their participation in the study.

Each participant received an identical packet including 16 vignettes of hypothetical scientific studies with corresponding graphs of the study's results. These vignettes included eight that described studies requiring some knowledge of cognitive psychology such as long-term memory as measured by the digit span task (see below).

An developmental psychologist was interested in how people's short-term memory changes as they get older. He gave children and young adults lists of numbers that they were asked to recall immediately following presentation. The most numbers that a person can remember is referred to as their "digit span." The researcher calculated the average digit span for each age group. On the right are results of his study. Please disucss the study.



A scientist at the sleep institute was interested in how many hours people slept a night as a function of age, so he surveyed a large number of people of different ages. On the right are his results. Please discuss the study.



The other eight vignettes/graphs described studies on general knowledge topics such as sleep (see above). See Appendix A for the full 16 vignettes.

Each vignette was followed only by the request to "discuss the study." Only one vignette and its corresponding graph was presented per page, each printed small enough that participants had plenty of room to write a response of up to a paragraph in length. Undergraduate students had only an hour to complete their packets in a laboratory setting, whereas graduate students were able to take the packet home to complete it in their free time.

We coded the participants' responses to the 16 graphs by categorizing statements in terms of how the participants described the data and the kinds of commentary they made about the data, the study or its results. The first dimension captured the participants' description of what they read and saw and the second dimension captured their criticisms or explanations of the studies and data presented in the vignettes. Two coders worked independently to code participants' responses. Cohen's kappa reliability tests were conducted on 20% of the coded data, and inter-rater reliability was .75. The coding scheme that I used is presented below:

Dimension 1. Sophisticated Descriptions of Data. Participants were given a "1" in this category if they provided a description of quantitative relationships between the variables. These descriptions included quantitative terminology such as "linear", "exponential", "more and more", "gradual", "slight", "inverted U", "steady increase/decrease", "sharpest increase/decrease", "varying rates", "consistent" or "levels off" or they included percentages or gave a description of the graph shape in a way that allowed the reader to visualize the graph at least approximately without looking at it.

Dimension 2. Explanation and Criticism of the Study and/or its Data. This dimension of my coding scheme categorized the types of commentary participants included in their responses. For each category within the axis, participants received a "0" if they did not include the category of commentary in their response and a "1" if they did include that category of commentary in their response. Participant explanation and criticism was split into the following categories:

A. No explanation or criticism- The participant failed to critique or provide comment on the study or results, i.e. no part of his/her statement fell under any of the categories that follow. If participants received a "1" in this category, they automatically received a "0" in the other interpretation categories.

B. Explanation- The participant identified an causal mechanism that he/she believed explained the pattern of results presented in the graph.

> E.g. "It makes sense that digit span should increase with age because people have had a longer time to practice working-memory tasks in their daily life compared to younger children who have had less need for these abilities."

C. Evaluation- A participant

a) included comments or criticism of the experimental design

> E.g. "This study should have been done longitudinally to account for the possibility that there is a generational difference due purely to cultural differences", "This experiment is flawed because it doesn't include individuals from all parts of the country or account for the difference relative value of a dollar in different areas"

b) made an evaluation of results by indicating that the results fit with prior knowledge, that they don't trust the results or by criticizing the results based on some prior knowledge.

> E.g. "That seems about right. It makes sense that older people would be worse drivers", "There's no way this is right. I am almost certain that baby's [*sic*]can't talk this soon."

c) asked questions by asking for further info, asking how the results could be explained, asking if a construct is an important one, or asking questions/making comments about statistical analysis leading to the results.

E.g. “Can a two year-old really do that?”, “How can you explain something like that?”, “Were the data split into separate groups before running analysis? What kind of comparison was done exactly?”

Results

For a full table of descriptive statistics from this study, see Table 1. In accordance with predictions, psychology graduate students provided the most sophisticated, quantitative description of the research findings, with undergraduates and history graduate students providing the least sophisticated descriptions (see Figure 1). For example, one social psychology graduate student provided a clear, quantitative description of a study about the relationship between vocabulary size and age: “Appears to be a linear relationship—increased age=increased vocabulary, although it plateaus, so resembles a power function more closely—highly accelerated vocabulary expansion at first, then slows down towards older age.” In contrast, an undergraduate student provided a much less sophisticated description of a study about the relationship between age and hours slept per night and age that doesn’t give a clear picture of the research finding: “From the age of 20, most people get less sleep.” Indeed, results from a one-way ANOVA indicated significant group differences on the description variable $F(4, 76)=4.42$, $p<.01$, $h_p^2=.189$.

Additionally, graduate students provided more evaluations of the data than did undergraduate students, with history and mechanical engineering graduate students providing the most evaluation (see Figure 2). For example, one history student questioned the definition of long-term memory as well as the effect size in a study examining the relationship between age and long-term memory for lists: “Isn’t remembering lists short-term memory? Can 1 week classify as long-term? Once again, the drop in # of words-8 to 5-doesn’t seem terribly significant.” This type of criticism, especially of definition and operationalization of variables, was typical of history students. Again, results from a one-way ANOVA indicated that these group differences on the evaluation variable were significant: $F(4, 76)=3.89$, $p<.01$, $h_p^2=.170$.

Undergraduates provided significantly more explanations of the data than did graduate students ($F(4, 76)=2.82$, $p<.05$, $h_p^2=.129$; see Figure 3). Examples such as the following explanation of the results of the age-vocabulary size study were typical of undergraduate responses: “As you get older you are exposed to more vocabulary because of surroundings. When you are younger you are exposed to different environments are less knowledgeable.” Graduate students were much less likely to provide explanations of the research findings as part of their interpretations.

We also conducted an analysis comparing people’s performance on the cognitive psych vs. general knowledge vignettes (see Figures 4, 5, & 6 for graphs of these results). For the undergraduate group, the graph content made a difference on all three dimensions of the coding scheme. Undergraduates provided significantly more sophisticated descriptions of the cognitive psych content vignettes as compared to the general knowledge vignettes ($F(1, 45)=18.18$, $p<.001$, $h_p^2=.288$. Undergraduates also provided significantly more evaluations of the cognitive psych content vignettes **(**$F(1, 45)=4.62$, $p<.05$, $h_p^2=.093$). However, the undergraduates provided significantly more explanations of the general knowledge vignettes **(**$F(1, 45)=16.43$, $p<.001$, $h_p^2=.267$). For example, the following are the same undergraduate students' responses to a cognitive psych vignette and a general knowledge vignette. The cognitive psych vignette dealt with the relationship between visual acuity and age: "OK! The graph is telling us that as you get older, your visual acuity reduces." The student simply provides a description of the research findings from this vignette. On the other hand, in response to a vignette dealing with the relationship between net worth and age the student wrote: "Well, hopefully this makes sense because if you have more than $20,000 in your account saved up by 20, you must really have been thinking ahead." In this response, the participant relates the research findings to his/her own beliefs and experience.

There were no significant differences in the way that cognitive psychology graduate students interpreted the cognitive psych vs. general knowledge graphs. In other words, cognitive psychology students' responses to cognitive psychology studies and non-cognitive psychology studies were surprisingly similar. For example, a cognitive psychology graduate student's response to the study about long-term memory for words was: "Memory for words declines in a nearly linear (constant rate) pattern, except for a sharper drop between 50 & 55." This same student's response to the study about number of hours slept per night: "Number of hours slept per night declines nearly linearly with age. At a rate of about ¼ hour loss for every 5 years." Other graduate students' provided more descriptions of cognitive psych as compared to general knowledge graphs**(**$F(1, 25)=8.40$, $p<.01$, $h_p^2=.252$), but there were no differences in the number of evaluations and explanations that they provided.

Discussion

Psychology graduates students who we can assume had the greatest comfort with interpreting behavioral research results, provided the most sophisticated descriptions of the data. This finding is consistent with Lehman et al.'s 1998 finding that psychology graduate students possess advanced statistical and methodological reasoning skills. It is possible that both groups of psychology students in my study used their statistical and methodological reasoning skills to make sense of the data provided to them and to describe them in way that took into account quantitative relationships between variables. Perhaps it was their statistical and methodological skills that allowed the psychology graduate students to provide complex descriptions of the research findings and graphs in a way that the other groups were not able to do. It is also possible that, because psychology students were familiar with the types of behavioral data presented in the vignettes, they were more accepting of these data, and, therefore, tended to describe it instead of evaluating it. The other groups of graduate students, while they did not provide many complex descriptions of the data, still did provide evidence-based responses to the vignettes in the form of evaluations.

Mechanical engineering and history graduate students in my study were consistently critical of the experimental results, questioning everything from the operationalization of variables to the procedure used to gather data, whereas undergraduates in my study tended to accept the validity of the results and jump right to the level of explanation. Interpreted through the lens of Kuhn's research on evidence vs. explanation-based reasoning, these results indicate that graduate students were more likely than undergraduates to provide evidence-based discussions of research findings. This result must be interpreted carefully because of the different amounts of time that the two groups (undergraduates and graduate students) were given to complete the tasks. However, this result is consistent with previous research showing that higher levels of education are associated with greater attention to evidence as opposed to explanation (Brem& Rips, 2000, Kuhn, 2001). It is also consistent with findings that higher levels of education are associated with critical thinking dispositions and skills more generally (Jehng et al., 1993; King & Kitchener, 2002; Toplak & Stanovich, 2003).

Undergraduates more than any other group were affected by the content of the vignettes that they were interpreting. They provided fewer explanations and more descriptions and evaluations of the cognitive psych vignettes as compared to the general knowledge vignettes. Likely this is because they did not have any prior knowledge relating to the content of the cognitive psych vignettes and, therefore, were not able to provide explanations for these research findings. It's interesting that the other graduate students groups were not as affected by the vignette contents, providing similar interpretations of the cognitive psych vs. general knowledge vignettes (with the exception of the number of descriptions provided by the non-cognitive psych graduate students). Perhaps increased higher educational experience makes one less vulnerable to content effects on evidence-based reasoning. This could be because, as part of their higher educational training, students are often asked to review and interpret evidence in one form or another. The number of explanations (as compared to data-driven responses) provided by graduate students was low regardless of the content of the vignettes.

In summary, the results from study 1 demonstrate that higher education is associated with a greater emphasis on evaluation of data as well as more sophisticated descriptions of quantitative information. The goal of study 2 was to examine how individual differences in epistemic beliefs, thinking dispositions and general intelligence may influence individuals' interpretations of data and explain education differences. In study 1, flaws identified by participants were not systematically built into the vignettes/graphs, so I was not able to include a measure of flaw recognition. In study 2 I built systematic flaws into the vignette measure that I used, and coded for whether individuals' noticed the specific flaws built into the vignettes.

Chapter 3

Study 2

Introduction

For study 2, I collected data on a new measure of everyday scientific reasoning that requires participants to comment on vignettes describing research findings. These vignettes had a number of qualities that the graphs and vignettes from the first study did not have. Firstly, I wanted to give subjects a task that is as close to a ‘real world’ evidence interpretation task as possible, in order gauge how they might behave in an everyday scientific reasoning context such as reading and interpreting a newspaper article. Therefore, for this study I decided to base most of my vignettes on newspaper articles from popular publications like the *New York Times* and the *San Francisco Chronicle*. I identified newspaper articles that described research findings, and adapted the descriptions from these articles to suit my purposes. I focused the vignettes on general knowledge topics that would be familiar to most if not all subjects such as obesity and video games. I wanted people to have experience with the issues raised in the vignettes, so that they had the potential to relate the research findings to their prior knowledge and beliefs.

Secondly, I wanted to address the question of whether students would be more likely to attend to evidence if the quality of the evidence was notably lower than in vignettes to include systematic methodological flaws such as sampling bias and correlation-causation confusion (see description of task below for further details). These flaws served to highlight the need for critical evidence evaluation in my task. I selected them based on three main factors. Firstly, these flaws often are likely to appear in newspaper articles I encountered; therefore, they seem to be numerous in the 'real world' context I was interested in reproducing. Secondly, literature on the teaching of statistics and research methods indicates that people seem to struggle with reasoning about many of these issues (Hatfield & Faunce, 2006; Leshowitz, 1989). Finally, participants in study 1 commented on these types of flaws in their discussions of the vignettes/graphs. This was particularly true of the issue of operationalization of variables.

Finally, I provided participants with two sets of instructions for interpreting the vignettes that I created. At first, participants were presented with the eight vignettes and were told simple to "discuss the studies." Then they were presented with the eight vignettes again and were asked to "critically evaluate the studies." I included these two sets of instructions in order to distinguish between what people were disposed to do and what they were able to do. This design allowed me to determine both whether people's initial, spontaneous reactions to the vignettes were evidence or explanation-based as well as whether they were able to shift to evidence-based responses when explicitly told to do so.

Study 1 examined how people trained in different fields of graduate study (and early-stage undergraduates) engage in everyday scientific reasoning about research findings. It provided a rich picture of the processes that less and more advanced students use to interpret research findings, with the major finding being that graduate students' interpretations are more data-driven than undergraduates'. However, while the findings of study 1 do indicate that there are educationally-driven differences in evidence interpretation, I do not know why this is. In other words, what individual difference factors underlie the reasoning of less and more advanced students? Are beliefs about knowledge (epistemic beliefs) and thinking dispositions an underlying factor driving evidence vs. explanation-based reasoning in this context, as Deanna Kuhn and others' work suggests? Are other factors such as statistical and methodological reasoning important as well when one is interpreting research findings?

As part of her work on scientific reasoning and critical thinking, Deanna Kuhn and others have shown that people with more sophisticated epistemic beliefs perform better on tasks requiring coordination of theory and evidence (Kuhn, 2001). Additionally, other research has demonstrated a positive relationship between thinking dispositions (e.g. actively open-minded thinking and need for cognition) and critical thinking constructs such as belief-bias avoidance (Stanovich& West, 1997). However, there is a dearth of research examining the relationships between epistemic beliefs and thinking dispositions and evidence-based interpretation of research findings. The following studies allowed me to address these issues, and should help me to identify underlying factors mediating the relationship between higher educational training and evidence interpretation. Specifically, I predicted that sophisticated epistemic beliefs and thinking dispositions would be positively associated with evidence-based reasoning about research findings. Additionally, I predicted that those who engage in evidence-based reasoning would be more likely to recognize the flaws in the vignettes. I also predicted that recognition of the flaws in the vignettes would depend on the amount of statistical and methodological training people had received.

In study 2, I tested an entirely undergraduate population. My participants were undergraduate students at less and more advanced stages of their undergraduate careers. This design allowed me to determine whether the differences I observed in the evidence interpretation processes of early undergraduate and graduate students were also observable in an entirely undergraduate population. Based on the results of study 1, as well as previous research on undergraduate training and reasoning (Kuhn, 2001; Lehman &Nisbett, 1990; Toplak & Stanovich, 2003), I predicted that undergraduate training would be positively associated with evidence-based reasoning. Therefore, I predicted that more advanced undergraduate students would provide more evidence-based interpretations of research findings. I also predicted that that those with training in statistics and research methodology (psychology majors) would be most likely to recognize the statistical and methodological flaws in the vignettes.

As in study 1, my unique open-ended data collection and coding allowed me to examine the processes by which students interpreted research findings. Additionally, the methodological flaws in my experiment vignettes allowed me to determine whether evidence-based interpretations of research findings were associated with identification of problems with these research findings. Finally, the additional measures of statistical and methodological training, and epistemic belief and thinking dispositions allowed me a better understanding of what factors underlay the processes that students' used to interpret research findings.

Pilot Study

I first conducted a pilot study, and gave 40 Introductory Psychology students the newly adapted vignettes as well as measures of thinking dispositions (actively open-minded thinking and need for cognition). Findings indicated a positive relationship between evidence-based interpretations of the research findings in the vignettes and actively open-minded thinking. Based on the results of the pilot study, I developed a study, which, in addition to the vignettes and dispositional measures, also included measures of epistemic beliefs, and amount of statistical and methodological training. I decided to include a measure of epistemic beliefs in addition to my dispositional measures because Kuhn makes the argument that people's epistemic beliefs shape intellectual values and hence the disposition to utilize intellectual skills (Kuhn, 2001). Thinking dispositions act to link epistemic beliefs and abilities in the context of critical thinking. By including an epistemic belief measure in this study, I was able to measure all of these interrelated constructs. For this study I recruited an additional group of Introductory Psychology students as well as advanced psychology, history and engineering students. This allowed me to examine potential educational differences in interpretation of vignettes, similar to the ones observed in study 1.

Participants were from the Introductory Psychology subject pool at the University of Michigan. I collected data on 40 participants, asking them about their majors, the science coursework they had taken and whether they considered themselves to be "good scientific reasoners." All participants completed the following measures on the SurveyMonkey.com website. All participants provided open-ended discussions of 8 vignettes describing research findings in response to two sequential prompts. At the first prompt, participants were simply asked to "discuss the 8 studies." At the second, they were asked to "critically evaluate these studies: the way they were conducted, their findings and/or their conclusions." All participants received both prompts in the same order. The majority of the vignettes were based on descriptions of experimental results from newspaper articles from the *New York Times* and the *San Francisco Chronicle*. Additionally, I systematically included a methodological flaw in each of the vignettes. The four types of flaws I included were: unclear operationalization of variables, over-interpretation of small effect sizes, biased samples and misinterpretation of correlational data (i.e. correlation/causation confusion). Each of the four flaws was included in two of the vignettes. The following vignette contains one of the operationalization flaws. In this vignette, what is meant by "creative accomplishments" is not defined.

> A researcher interested in how creativity in scientific reasoning changes with age conducted a study. He asked colleagues of hundreds of randomly chosen scientists from across the United States to rank the mean number of creative scientific accomplishments that their colleagues had made over the past year. He calculated with average number of creative accomplishments as a function of age, and found that number of creative accomplishments increased from age 20-35 and then began to decline.

This next vignette contains an effect size flaw. The small effect size (3 points difference) is over-interpreted as a momentous and important finding.

> The eldest children in families tend to develop higher I.Q.s than their siblings, researchers are reporting today, in a large study (two papers) that could settle more than a half-century of scientific debate about the relationship between I.Q. and birth order. The average difference in I.Q. was three points higher in the eldest child than in the closest sibling. "I consider these two papers the most important publications to come out in this field in 70 years; it's a dream come true," said Frank J. Sulloway, a psychologist at the Institute of Personality and Social Research at the University of California, Berkeley.

This third example contains one of the sampling bias flaws. People asked about interest in politics in the study were already participating in political events, and were therefore not a representative sample of the population of Americans.

> A recent study shows that Americans are more interested in politics than was previously thought. Researchers approached people at events such as town meetings and city council meetings and surveyed them about their interest in political and their voting behavior. Seventy percent of those surveyed reported that they were planning to vote in upcoming local and national elections. More than half said that they regularly read articles in the newspaper about political issues.

This final example contains one of the correlation-causation flaws. In this example, the relationship between controlling mother and obese children is misinterpreted as a causal one.

> A study of 77 children aged 3 to 5 found that those with the most body fat had the most "controlling" mothers when it came to the amount of food eaten. "The more control the mother reported using over her child's eating, the less self-regulation the child displayed," Dr. Johnson and her co-author said.

For the full set of 8 vignettes used in the pilot study, see Appendix B.

The Actively Open-Minded Thinking scale is a thinking dispositional questionnaire designed to measure people's ability to engage in actively open-minded thinking (Stanovich& West, 1997; Sá, Kelley, Ho, &Stanovich, 2005).The actively open-minded thinking scale was composed for 41 items taken from different sources:10 items from a flexible thinking scale developed by Stanovich and West(1997); 8 items from the Openness-Values facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992); 9 items measuring dogmatism (Paulhus & Reid, 1991; Robinson, Shaver, &Wrightsman,1991; Troldahl & Powell, 1965); 3 items from the categorical thinking subscale of Epstein and Meier's (1989) constructive thinking inventory;9 items from the belief identification scale developed by Sa et al. (1999);2 items from a counterfactual thinking scale developed by Stanovich and West (1997). The response format is a 6-point Likert scale (6—agree strongly, 1--disagree strongly). Sample items include: “A person should always consider new possibilities,” and “There are two kinds of people in this world: those who are for the truth and those who are against the truth” (Reflected). See Appendix D for the full Actively Open-Minded Thinking scale.

We also included the Need for Cognition scale, another dispositional scale measuring the motivation to engage in challenging cognitive activities (Caccioppo & Petty, 1982). Participants were asked whether a series of statements are characteristic of them. The response format was a 5-point Likert scale (5---extremely characteristic, 1---extremely uncharacteristic). Sample items include: “I find satisfaction in deliberating hard and for long hours,” and “I only think as hard as I have to” (Reflected). See Appendix E for the full Need for Cognition scale. For both the AOT and NFC questionnaires, responses on reverse items were reversed and responses were summed to create one composite score for each participant.

We adapted the coding scheme used in study 1 to address the specific question of whether participants recognized methodological flaws in the vignettes. Additionally I coded whether the participants evaluated or critiqued the data and/or whether they provided an explanation for the pattern of results. I considered evaluations of the data to be evidence-based reasoning and explanations of the data to be explanation-based reasoning. This adapted coding scheme did not focus on participants’ description of the results because the results were already described in the vignettes (and are not accompanied by a graphical representation). Therefore, I determined that it was unlikely that participants would spend much time on description. The coding scheme that I used is presented below.

1. Explanation and Evaluation. This section of my coding scheme categorized the types of commentary participants included in their responses. For each category within the axis, participants received a “0” if they did not include the category of commentary in their response and a “1” if they did include that category of commentary in their response. This section is was divided into the following categories:

A. No explanation or criticism- The participant failed to critique or provide comment on the study or results, i.e. no part of his/her statement fell under any of the categories that follow. If participants received a “1” in this category, they automatically received a “0” in the other interpretation categories.

B. Evaluation/Criticism of Methodology- The participant

   a. Questioned the definition or operationalization of a concept or category like creativity (e.g. general health too broad of a category)

   b. Questioned the survey materials, scales used, experimental materials- (e.g. # of word lists)

   c. Questioned the experimental design/procedure

      e.g. Number of ss, design (longitudinal), sample –random?, age range etc.

      e.g. How was x calculated?

      e.g. How was data collected?

      e.g. Statistical criticisms

d. Questioned the interpretation of the results

e.g. The authors interpret the results as a causal effect (when it is really just correlational).

e.g. They authors over-interpret an effect (i.e. a small effect is blown out of proportion)

e.g. Contradiction of author's interpretation of results (i.e. maybe this could be interpreted differently….)

C. Recognition of major flaw- The participant recognized the major flaw (listed below).

a. **Vignette 1(obese children with controlling mothers):** It is a correlational study>>>>you cannot derive causation (i.e. just because obese children have more controlling mothers, doesn't mean that the mother over-control CAUSED them to be obese. The mother could also be controlling BECAUSE the children are obese).

b. **Vignette 2(creativity**): There is NO operational definition of creativity (i.e. how are the scientists supposed to know what he/she means by "creative accomplishments" when they rank their colleagues?)

c. **Vignette 3 (abuse):** There is NO operational definition of abuse (what do they mean when they say the have been "abused"?)

d. **Vignette 4 (runners):** It is a correlational study>>>>you cannot derive causation (i.e. just because these people run, it doesn't mean that their running CAUSED them to be physically fit. It could be diet or other factors).

e. **Vignette 5 (IQ):** This is a very small effect size. (i.e. There is only 3 points of difference on the IQ test, but they make it sound like a much bigger finding than it actually is).

f. **Vignette 6 (Video games):** This is an issue of sampling bias. (i.e. The kids tested are all the children of university professors; therefore, they are not a random sample of the population).

g. **Vignette 7 (Preschool):** This is a very small effect size. (i.e. There is only a 1 percent difference on the test, but they are making it sound like a much bigger finding than it actually is).

h. **Vignette 8 (Political participation):** This is an issue of sampling bias. (i.e. The people surveyed are already at political events; therefore, they are not a random sample of the population).

D. Explanation- The participant identified a causal mechanism that he/she believed explained the pattern of results presented in the vignettes based on own his/her beliefs, theories, and experiences.

Descriptive statistics for the Need for Cognition Scale, the Actively Open-Minded Thinking Questionnaire and the vignette coding are presented in Table 2. One can see from this that people were more likely to explain the research findings (for an average of over 5.5 out of 8 vignettes) than they were to evaluate them (for an average of 3 out of 8 vignettes). Additionally, people only recognized the major flaw for an average of 1.5 out of 8 vignettes. When participants were told directly to critically evaluate the studies, the average number of studies evaluated jumped to 4 and the number of flaw recognized jumped to 2; however, people still provided explanations for approximately 6 vignettes.

Analyses conducted on 40 participants' data revealed a significant positive correlation between scores on actively open-minded thinking and number of vignettes that were evaluated $r=.383$, $p<.05$ as well as the number of times participants recognized the major flaw $r=.526$, $p<.01$. Not surprisingly, evaluation of the vignettes and recognition of the major flaw were strongly positively correlated $r=.755$, $p<.01$. The Need for Cognition scale was not significantly correlated with any of the other measures.

Interestingly, evaluation of the vignettes was positively correlated with explanation of the vignettes $r=.488$, $p<.01$, suggesting that some people both evaluated and explained the research findings. Additionally, the number of explanations provided was positively correlated with the Actively Open-Minded Thinking questionnaire, although this correlation did not reach significance $r=.294$, $p>.05$).

These results indicated that people's responses to the vignettes were successfully coded into evidence and explanation-based categories, with this group tending to provide more explanations than evaluations of the research findings. The number of flaws recognized by the participants was relatively low (1.5 out of 8), and I altered the vignettes before conducting study 2 to make the flaws more conspicuous. The positive correlation between actively open-minded thinking and the number of evaluations provided and flaw recognized was predicted; however, the positive correlation between actively open-minded thinking and the number of explanations provided was surprising. Additionally, I was surprised to find that need for cognition was not significantly correlated with any of the vignette measures. These issues were explored further in study 2.

Study 2 Methodology

Study 2 allowed me to examine the relationship between evidence vs. explanation-based interpretation of experimental results, epistemic beliefs and thinking dispositions, statistical and methodological training and higher education. In this study, I included all measures from the pilot study in addition to a measure of epistemic beliefs (before I only measured dispositions). I decided to include a measure of epistemic beliefs in addition to my dispositional measures because Kuhn makes the argument that people's epistemic beliefs shape intellectual values and hence the disposition to utilize intellectual skills (Kuhn, 2001). For example, if one believes that there can be multiple valid theories that must be compared with one another, one will be disposed to use one's intellectual skills to seek out and evaluate evidence related to these theories. Thinking dispositions act to link epistemic beliefs and abilities in the context of critical thinking. By including an epistemic belief measure in this study, I was able to measure all of these interrelated constructs. Additionally, I tested both Introductory Psychology students and undergraduate students trained in psychology, history and engineering.

Study 2 involved freshman participants from the Introductory Psychology subject pool at the University of Michigan as well as more advanced (sophomore, junior and senior) undergraduate students in the University of Michigan departments of Psychology, Engineering and History. Frequencies broken down by year in college and major are presented in Table 3. This mirrors the cross-sectional design used in study 1 with advanced undergraduates instead of graduate students.

Based on the small number of flaws recognized in the pilot study (1.5 out of 8), I changed the wording of the vignettes to make the flaws even more apparent. Below is an example of a vignette from the pilot study, and then again in its revised version as it was used in study 2.

> A study of 77 children aged 3 to 5 found that those with the most body fat had the most "controlling" mothers when it came to the amount of food eaten. "The more control the mother reported using over her child's eating, the less self-regulation the child displayed," Dr. Johnson and her co-author said.
>
> As part of a recent study, researchers measured children's body fat and surveyed mothers about the amount of control they exert over their children's eating. The results of this study, conducted with 77 children aged 3 to 5, found that those with the most body fat had the most "controlling" mothers when it came to the amount of food eaten. This shows that when mothers exert more control over their children's eating, the children display less self-control, researchers said.

The final line in the second version of the vignette was changed so that it would emphasize that the correlational relationship in this study was incorrectly interpreted as a causal relationship. Additionally, the name "Dr. Johnson" was removed from the vignette to avoid potential effect of this source information on people's interpretation of the study (Petty, Priester, & Brinol, 2002).

As in the pilot study, participants completed all measures on the Survey Monkey website, beginning with the vignettes ("discuss studies" condition followed by "critically evaluate" condition), followed by the epistemic belief and thinking disposition questionnaires and the background/follow-up questions. Additionally, to determine whether the web context had a significant effect on participants' performance on the tasks, I also tested 15 participants in a lab setting. There were no significant differences between the performance of lab and non-lab participants on any of my measures (see Table 4 for non-significant ANOVA vignettes results); thus, I report data combined. The measures that were used in this study were the same as those described for the pilot study, with a few notable changes. Firstly, in addition to asking about college major and coursework (including their statistics and research methods courses), I also asked students to report on their beliefs about the research topics discussed in the vignettes, their knowledge of the specific types of methodological flaws included in the vignettes, their interest in media outlets such as newspaper articles and their SAT or ACT scores (see Appendix G for full list of questions). Finally, I included a measure of epistemic beliefs in addition to my actively open-minded thinking and need for cognition measures. The measure I used to measure epistemic beliefs was the Epistemic Belief Inventory (EBI), developed by Schraw, Bendixen and Dunkle (2002). It consists of 28 items to which participants respond using a five-point Likert scale in which 1 corresponds to "strongly disagree" and 5 corresponds to "strongly agree." Sample items include: "most things worth knowing are easy to understand" and "what is true is a matter of opinion." The full scale is included in Appendix F. The coding scheme used to code the vignettes was the same as the one used in the pilot study.

Results

See Table 5 for a full table of descriptive statistics for all tasks used in study 2. Vignette data was coded for evidence-based (evaluations of the data) vs. explanation-based (explanations of the data) reasoning as well as for recognition of the major flaws in the vignettes. Coding and analysis of vignette data indicated that my data did conform to an evaluation vs. explanation model of scientific reasoning. People's responses fit easily into my evaluation and explanation coding categories and 20% inter-rater reliability was good (Cohen's kappa=.7).

When presented with my 8 vignettes (see Appendix B), and told simply to "Discuss the studies," people provided evaluative statements for an average of 3.45 out of

8 vignettes (SD=2.35), explanatory statements for an average of 2.88 vignettes (SD=2.01), and they recognized the major flaw in an average of 1.6 vignettes (SD=1.47). See Figure 7 for a graph of these results. For example, one participant provided the following evaluative statement in response to the vignette describing the study on political interest:

> This research is clearly not very accurate as the people surveyed were already participating in politics. Of course they are going to have an interest in politics! To make this study more accurate, a random sample would have to be done at events that are not political in nature.

In this example, one can see that this participant also recognized the major flaw in the vignette: sampling bias. Another participant provided the following explanatory statement in response to the vignette describing the study on controlling mothers and obese children: “Mothers think they can get their child to be skinnier by controlling them when really it just makes a child want to disobey them.” In this example, the participant accepted the research findings and immediately attempted to explain them. As in the pilot study, when participants were specifically told to “critically evaluate the studies,” they were more likely to evaluate than when they were simply told to “discuss the studies” (M=5.17, SD=2.16). They were also slightly more likely to recognize the major flaw in the study under these circumstances (M=2.13, SD=1.48). However, interestingly, when told to critically evaluate the studies, people were also more like to explain the study (M=3.24, SD=2.37).

When I examined vignettes with different types of flaws built into them, I discovered some interesting patterns. People were most likely to evaluate and recognize flaws in the vignettes with built in sampling bias errors. They were least likely to recognize flaws in the vignettes with built in over-interpretations of small effect sizes and correlation-causation misinterpretation (see Figure 8). This is in line with previous research suggesting that people have particular difficulty distinguishing between correlation and causation (Hatfield and Faunce, 2006; Zimmerman, 2005). However, there is a dearth of research comparing people's recognition of different types of statistical and methodological flaws, so it is difficult to interpret these differences.

At the end of the study, I asked people about their beliefs about and experiences with the topics presented in the vignettes. As would be expected, people tended to provide evaluations for a higher percentage of the vignettes with results that contradicted their prior beliefs and experiences (41%) than the vignettes with results that in line with their beliefs and experiences (28%). The same was true for recognition of flaws, with people recognizing flaws in 22% of the vignettes that contradicted their beliefs and experiences and 8% of the vignettes in line with their beliefs and experiences. This result fits with previous research showing that people are more critical of evidence that goes against their prior beliefs (Lord, Ross & Lepper, 1979).

We conducted ANOVA analyses to determine whether there were group differences in interpretation of the vignettes, as well as in performance on my survey measures. I compared the performance of freshman, sophomore, junior and senior undergraduate students as well as the performance of advanced undergraduates with different types of educational training (psychology, history, engineering).

An analysis of variance revealed a significant effect of year in college on students' interpretation of the vignettes (see Figure 9; Table 6). In the 'discuss this studies' condition, there were significant differences in the number of evaluations provided by students in different years in college $F(3, 264)=11.23$, $p<.001$, $h_p^2=.113$, with freshman providing the fewest evaluative statements and seniors providing the most. Recognition of major flaws in the vignettes exhibited the same significant pattern, with students with more education recognizing more and more flaws $F(3, 264)=9.81$, $p<.001$, $h_p^2=.100$. The following example is a response provided by a junior mechanical engineering major when asked to "discuss" the vignette describing the creativity study:

> This study is very much based on speculation by those interviewed. Especially at question is the definition of the variable. What constitutes a "creative scientific achievement;" a survey done by one, impartial scientist or panel would be more effective at accurately and unbiased study.

The number of explanations showed the opposite pattern, with freshman and sophomores providing the most explanations and juniors and seniors providing the fewest, although this pattern was not significant $F(3, 264)=1.47$, $p=.224$, $h_p^2=.016$. The following is the response of a freshman undeclared major when asked to "discuss" the study on creativity:

> This makes sense because the age of 20-35 range is a time where most adults prosper. They are no longer kids and are able to make their own real accomplishments. After 35 people tend to have a steady job and start families. The exciting, risk taking, accomplishing part of life slowly turns into a routine.

This freshman student accepts the research findings described in the vignette and immediately attempts to make sense of them in terms of his/her prior knowledge and beliefs.

When covariates were included in the models described above, this ANCOVA analysis showed a significant effect of dispositional and percentile measures as covariates. In the corrected model, there were still significant differences in the number of evaluations provided by students with different amounts of education ($F(6, 252)=7.43$, $p<.001$, $h_p^2=.081$), although this effect was lessened by the introduction of covariates. Score on the Need for Cognition questionnaire and percentile on the SAT/ACT also significantly and marginally significant affected the number of evaluations provided in this condition (NFC: $F(6, 252)=5.61$, $p<.05$, $h_p^2=.022$; Percentile: $F(6, 252)=3.51$, $p=.062$, $h_p^2=.014$). There were also still significant differences in the number of times major flaws were recognized by students with different amounts of education ($F(6, 252)=6.00$, $p=.001$, $h_p^2=.067$), although this effect was also lessened by the introduction of covariates. Score on the Actively Open-Minded Thinking questionnaire also significantly affected the number of flaws recognized ($F(6, 252)=6.34$, $p<.05$, $h_p^2=.025$). There were no significant covariate effects on the number of explanations people provided.

An analysis of variance revealed a significant effect of major on students' responses to the vignettes (see Figure 10; Table 7). In the 'discuss the studies' condition, there were significant differences in the number of evaluations provided by students with different majors $F(4, 263)=7.70$, $p<.001$, $h_p^2=.105$, with history majors providing the most evaluative statements, followed by engineering majors. Psychology and undeclared majors provided the fewest evaluative statements. For example, one history major's response to the research finding that children enjoy educational video games as much as they enjoy non-educational video games was:

> Interesting data, but I am inclined to argue that the study doesn't necessarily indicate anything about children as a whole. A fuller study of children of every social or economic class would be needed before they can claim that all children enjoy educational and non-educational games alike.

This student not only evaluates the methodology of the study described, but recognized the major flaw in the vignette (sampling bias). Recognition of major flaws in the vignettes exhibited the same significant pattern ($F(4, 263)=3.76$, $p<.01$, $h_p^2=.054$), with history and engineering majors recognizing the most flaws. The number of explanations did not display a significant effect of major.

Similar patterns of group differences emerged for my surveys, measuring thinking dispositions and beliefs (see Tables 8 and 9) as well as self-reported interest in and knowledge of scientific content and methodology (see Tables 10 and 11), with seniors exhibiting the most sophisticated thinking beliefs and dispositions as well as the strongest interest in and knowledge of scientific content followed by juniors, sophomores and freshman (see Table 12). Interestingly, engineering majors had the highest need for cognition scores, followed closely by the history majors. History majors had by far the highest actively open-minded thinking questionnaire scores, followed by the engineering majors (see Table 9). This is interesting considering that history and engineering majors also provided the most evidence-based responses to the research findings and recognized the most flaws in the vignettes.

When covariates were included in the models described above, this ANCOVA analysis showed a significant effect of dispositional and percentile measures as covariates. In the corrected model, there were still significant differences in the number of evaluations provided by students with different majors ($F(7, 251)=5.22$, $p<.001$, $h_p^2=.077$), although this effect was lessened by the introduction of covariates. Score on the Need for Cognition questionnaire and percentile on the SAT/ACT also significantly and marginally significant affected the number of evaluations provided in this condition (NFC: $F(7, 251)=5.40$, $p<.05$, $h_p^2=.021$; Percentile: $F(7, 251)=3.35$, $p=.068$, $h_p^2=.013$). In the ANCOVA model, there were now only marginally significant differences in the number of times major flaws were recognized by students with different amounts of education (F(7, 251)=2.06, p=.086, h_p^2=.032). Score on the Actively Open-Minded Thinking questionnaire and percentile on the SAT/ACT also significantly and marginally significantly affected the number of flaws recognized (AOT: F(7, 251)=7.01, p<.01, h_p^2=.027; Percentile: F(7, 251)=2.92, p=.89, h_p^2=.011). There were no significant covariate effects on the number of explanations people provided.

We also conducted a 2 x 2 mixed ANOVA on number of evaluations, number of recognitions and number of explanations and found no significant interactions between condition (discuss vs. critically evaluate) and amount of education or between condition and major of interest (see Table 16). This indicates that, under the "critically evaluate" condition (as opposed to the "discuss this study" condition), participants with different amounts of education and different majors increased the number of evaluations, recognitions and explanations they provided by similar amounts.

We conducted correlational analyses to help me better understand the relationship between my survey measures and students' interpretation of the vignettes. A full table of non-vignette data correlations is presented in Table 14. As expected, correlational analyses revealed that the Actively Open-Minded Thinking Questionnaire, the Epistemic Beliefs Inventory and the Need for Cognition Scale were all significantly positively related to each other (AOT & EBI: *r=.323, p<.01*; AOT & NFC: *r=.400, p<.01*) . Actively open-minded thinking and need for cognition were also positively related to SAT/ACT percentile (AOT: *r=.185, p<.01*; NFC: *r=.261, p<.01*). There were also some significant positive relationships between the dispositional measures (AOT and NFC) and people's self-reported interest in and critical reading of scientific studies and their self-reported familiarity with the statistical flaws built into my vignettes.

Analyses with the vignette data (see Table 15), as expected, revealed that the number of evaluations that people provided was strongly positively related to the number of major flaws they recognized in both the “discuss the studies” and “critically evaluate the studies” conditions (DS: $r=.790$, $p<.01$; CE: $r=.620$, $p<.01$). However, unlike in the pilot study, the number of evaluations people provided was negatively related to the number of explanations they provided ($r= -338$, $p<.01$) in the “discuss the studies” condition, indicating that people either evaluated or explained the vignettes but not both. The number of explanations that people provided was also negatively related to the number of flaws that people recognized ($r=-.280$, $p<.01$) in the “discuss the studies” condition.

Additionally, the dispositional measures (the Actively Open-Minded Thinking Questionnaire and the Need for Cognition Scale) were positively related to the number of evaluative statements the people provided in response to the vignettes (AOT: $r=.209$, $p<.01$; *NFC:* $r=.269$, $p<.01$) and to the number of times people recognized major flaws in the vignettes (AOT: $r=.271$, $p<.01$; NFC: $r=.238$, $p<.01$) in the “discuss the studies” condition. Interestingly, these correlations were also positive and significant in the “critically evaluate” condition. The dispositional measures were not related to the number of explanations that people provided. The Epistemic Belief Inventory was only significantly positively related to the number of evaluations provided in the “discuss the studies” condition.

The number of statistical courses that a student had taken was also positively related to the number of evaluative statements provided in response to the vignettes ($r=.133$, $p<.05$) and to the number of flaws recognized ($r=.126$, $p<.05$) in the “discuss the studies" condition. This was also true of the "critically evaluate the studies" condition. However, the number of statistical course taken was negatively related to the number of explanations provided in the "discuss the studies" condition, although this result was not significant ($r=-.102, p>.05$).

We used regression analyses to examine whether evaluation of the vignette tasks and recognition of major flaws was predicted by general intelligence, thinking beliefs and dispositions and/or amount of statistics and methodological training. I created 4 regression models: two with the number of evaluations as the dependent variable (one for the "discuss the studies" condition and one for the "critically evaluate the studies" condition), and two for the number of major flaws recognized (one for the "discuss the studies" condition and one for the "critically evaluate the studies" condition). For each of the models, I entered percentile on the ACT/SAT, number of statistics/methodology courses taken, and scores on the EBQ, AOT and NFC questionnaires as independent variables in the model.

The first model, predicting number of evaluations provided when given the "discuss the studies" instructions explained a significant proportion of the variance, R^2 = .113, $F(5, 258) = 6.46$, $p< .001$, with percentile on the SAT/ACT and need for cognition both significantly predicting number of evaluations (Percentile: b=.124, $t(258)=1.99$, $p<.05$; NFC: b=.151, $t(258)=2.21$, $p<.05$).

The second model, predicting number of evaluations provided when given the "critically evaluate" instructions explained a significant proportion of the variance, R^2 = .091, $F(5, 258) =5.09$, $p< .001$, with number of statistics and methodology courses taken and need for cognition both significantly predicting number of evaluations (Stats/Methods Courses: b=.156, $t(258)=2.56$, $p<.01$; NFC: b=.170, $t(258)=2.45$, $p<.05$).

The third model, predicting number of flaws recognized when given the "discuss the studies" instruction explained a significant proportion of the variance, $R^2 = .112$, $F(5, 258) = 6.40$, $p< .001$, with actively open-minded thinking significantly predicting number of flaws recognized (AOT: b=.204, $t(258)=3.05$, $p<.01$).

The fourth model, predicting number of flaws recognized when given the "critically evaluate" instruction explained a significant proportion of the variance, $R^2 = .143$, $F(5, 258) = 8.45$, $p< .001$, with percentile on the SAT/ACT, number of statistics and methodology courses taken and actively open-minded thinking all significantly predicting number of flaws recognized (Percentile: b=.118, $t(258)=1.93$, $P=.05$; Stats/Methods Courses: b=.115, $t(258)=1.94$, $p<.05$; AOT: b=.270, $t(258)=4.11$, $p<.001$).

Discussion

The results of this study give us insight into the processes that people use to evaluate research findings, as well as some understanding of why they engage in these processes. Ease of coding and strong inter-rater reliability indicates that Deanna Kuhn's model of theory-evidence coordination can successfully be applied in this context. The participants' responses easily fell into the categories evidence-based and explanation-based. Despite the fact that the evidence presented to them contained major flaws, many participants still provided explanation-based responses to this evidence. However, participants were able to shift and provide more evidence-based responses when directly instructed to do so (in the "critically evaluate" condition).

We also found educational differences in the ways that people interpreted my flawed research findings vignettes. As in previous studies on critical thinking (Jehng et al., 1993; King & Kitchener, 2002; Toplak & Stanovich, 2003), I found that those with more educational experience (juniors, seniors) provided more critical, evidence-based responses to the vignettes while those with less educational experience (freshman, sophomores) provided more theory-driven and explanation-based responses. ANCOVA results indicated that the educational differences in number of evaluations provided and flaw recognized were perhaps mediated by both ability and dispositional factors. These results are in line with previous findings that higher educational experience is related to development of sophisticated epistemic beliefs and critical thinking skills such as belief-bias avoidance.

One limitation of this study is that the type of compensation (credit or payment) was confounded with the year in college variable. As is always the case, the students in Introductory Psychology were primarily freshman and sophomores, whereas the students I recruited from the Psychology, History and Engineering Departments were primarily juniors and seniors. This meant that more of the participants who received credit were freshman and sophomores and more of the participants who received payment were juniors and seniors. However, the type of compensation seemed to have little effect on people's performance on the vignette tasks. When examined by year in college, compensation did not have a significant effect on any of the vignette coding dimensions except for on the number of explanations provided by sophomores in the "critically evaluate" condition ($F(1, 76)=4.99$, $p<.05$, $h_p^2=.062$)

Additionally, as in Lehman & Nisbett's 1990 research on effects of undergraduate training on reasoning, I found differences in the ways students from different majors interpreted research findings. History and engineering students tended to be the most evaluative, offer a wide range of criticisms of the research findings presented in the vignettes and recognizing the most major flaws of any of the groups. Interestingly, psychology students were not as critical as history and engineering students, and did not recognize as many major flaws in the vignettes. This is in line with my findings from study 1 in which history and mechanical engineering tended to provide the most evaluations of vignettes. Perhaps this is because the psychology students were the most familiar with the types of findings presented in the vignettes (behavioral research findings), and therefore tended to accept them as valid. However, history and engineering students also had higher scores on the thinking disposition and ability measures, and ANCOVA results indicated that the educational differences in number of evaluations provided and flaw recognized were perhaps mediated by these ability and dispositional factors.

We also found that many of predictions about the relationships between variables of interest were born out. The dispositional measures were related to the number of evaluations provided and the number of flaws recognized, but they were not related to the number of explanations that people provided. This indicates that, as Deanna Kuhn's research suggests, critical thinking dispositions such as actively open-minded thinking and need for cognition are associated with evidence-based and not explanation-based reasoning.

My regression analyses revealed interesting patterns of results. The four models that I tested all predicted a significant proportion of the variance. Interestingly, the number of evaluations provided in the “discuss the studies” condition was predicted by percentile on the SAT/ACT and need for cognition, whereas number of evaluations provided in the critically evaluate condition was predicted by the number of statistics and methodology classes taken and need for cognition. This indicates that, when given freedom to interpret the research findings, those with a combination of general intelligence and the motivation to use that intelligence, will be both disposed and able to carefully evaluate what they are reading. However, when explicitly told to critically evaluate the research findings, it is those with the specific statistical and methodological experiences who are able to do so. This is largely in line with my original predictions that statistical and methodological skills would be most important when participants were directly told to critically evaluate vignettes.

The number of flaws recognized when given the “discuss the studies” instructions was significantly predicted by actively open-minded thinking, whereas the number of flaws recognized in the “critically evaluate” condition was predicted by actively open-minded thinking, percentile and the number of statistics and methodology courses taken. This is interesting because it indicates that it is a dispositional factor that predicts people’s tendency to seek out and perform the task at hand, even when not explicitly told to do so.

Chapter 4

General Discussion

The current work examines the processes that people use to interpret research findings and helps us to understand why they engage in these processes. It makes a contribution to the literature in two important ways: by applying Deanna Kuhn's theory-evidence coordination in a new and important context, and by providing a better understanding of experiential, dispositional and ability factors underlying theory-evidence coordination. One of the greatest strengths of this work is its innovative methodology that allows for analysis of open-ended, authentic data. Creating instruments and coding schemes that examine the process by which people interpret research findings as well as applying previously validated instruments to understand why people engage in this process are important steps in understanding what factors allow people to become careful consumers of evidence.

I do not attempt to argue against the importance of explaining evidence and incorporating it into one's existing body of knowledge and beliefs (Zimmerman, 2000). There are many circumstances in which theory-driven reasoning is appropriate (Meehl, 2002). For example, Koslowski (1996) argues that scientist rely on theory or mechanism to decide which of the many covariations in the world are likely to be causal, and that this is a scientifically legitimate way to reason. I agree with this argument; however, I would argue there are certain circumstances that demand that evidence be considered carefully before one moves on to the level of explanation. This work is an attempt to examine the ways in which people operate in such situations as well as the factors that underlie their approach to such evidence.

Looking across the two studies, my results indicate that people's interpretations of experimental results can be successfully coded using models such as Kuhn's theory-evidence coordination model into evidence-based and explanation-based reasoning. These results are a promising indication that authentic, open-ended data can be successfully managed and used to study critical thinking.

These results also provide additional support for Deanna Kuhn's model of theory-evidence coordination. In reasoning processes that conform to the theory-evidence coordination model, evidence figures heavily, and must be carefully evaluated in order to determine its quality. Theories are developed in tandem with evidence, and it is the theory associated with the higher quality and most consistent evidence that is selected. On the other hand, in reasoning processes that conform to a satisficing model, evidence does not play a large role and is largely disregarded, especially when it is inconsistent with previously existing theories. Plausible explanations are sufficient to dictate the validity of a theory and alternative theories are not considered. My findings lend support to the existence of these two types of reasoning. In my studies, participants' responses were easily and successfully coded into evidence and explanation-based responses. Additionally, participants' evidence-based responses were related to and predicted by some of the educational, belief and dispositional factors that Kuhn discusses in her work (Kuhn, 2001). Therefore, my work lends support to Kuhn's model and extends its scope to a new and important domain of study: people's open-ended interpretation of research findings.

I found that those with more higher educational training tended to provide more evidence-based interpretations of research findings. In both of my studies, those at more advanced stages their higher education (graduate students in study 1, advanced undergraduates in study 2) provided more evaluations of the research findings presented to them. When I examined educational differences by major/field of study, I found that psychology students tended to provide explanations and descriptions of the research findings as opposed to evaluations of them. On the other hand, history and engineering students were more likely to evaluate (and often criticize) the research findings. This distinction could be a product of the fact that psychologists are more familiar and therefore perhaps more comfortable with the kinds of behavioral research findings presented in the vignettes. History and engineering student are less familiar and therefore perhaps more wary of the methodologies used in the studies described.

These differences could also be driven by differences on my thinking disposition and ability measures. In study 2 I found that history and engineering majors tended to have higher scores on actively open-minded thinking (especially history majors), need for cognition (especially engineering majors) and SAT/ACT percentile (especially engineering majors) than psychologists. It could be that these factors contributed to the larger number of evaluations provided by these groups. Additionally, ANCOVA analyses revealed significant effects of dispositional and ability measures as covariates. That interpretation would be consistent with previous findings that those with sophisticated thinking dispositions perform well on critical thinking tasks (Klaczynski & Robinson, 2000; Stanovich & West, 1997).

These findings extend previous work looking at higher educational effects on critical thinking constructs such as epistemic beliefs, thinking dispositions, belief-bias avoidance and theory-evidence coordination (Jehng et al., 1993; King & Kitchener, 2002, Kuhn, 2001; Toplak & Stanovich, 2003). In my studies, as in other studies, I found that students with more higher educational training tended to have more sophisticated epistemic beliefs and thinking dispositions, and tended to engage in more critical, evidence-based reasoning. Additionally, my findings extend the work of Lehman, Lempert & Nisbett (1988) and Lehman & Nisbett (1990), which examined the effects of discipline-specific graduate and undergraduate education on a range of reasoning skills. Findings from my studies show discipline-specific differences in the ways that graduate and undergraduate students approach the task of interpreting research findings. Psychology students in both my graduate and undergraduate studies were more accepting of the research findings as compared with other graduate and advanced undergraduate groups, most likely because of their familiarity with the types of findings being presented. Interestingly, however, these students' statistical and methodological training did not help them to recognize the major flaws that I built into my studies. Further research is needed in order to better understand the performance of my discipline-specific groups.

Findings from study 2 reveal significant relationships between many of the factors of interest. As expected, factors such as general intelligence, thinking dispositions, and number of statistics courses taken were positively related to evidence-based reasoning (and negatively related to explanation-based reasoning). Additionally, I created a number of predictive models that provide a better picture of what factors underlie evidence-based reasoning and flaw recognition. It is clear that, while some less malleable factors such as general intelligence are important predictors of people's evidence-based reasoning, there are other predictive factors that are experiential and/or possible to train. For example, I find that actively open-minded thinking was a strong predictor of the number of vignette flaws a person recognized. This is a disposition that is closely related to epistemic beliefs and is very likely affected by higher educational experience (Jehng et al., 1993). Additionally, the number of statistics and methodology courses people had taken was predictive of the number of evaluations they provided and the number of flaws they recognized when given the instructions to critically evaluate the vignettes. Again, this is a good example of experiential factors that contribute to a person's ability to think carefully and critically about research findings.

Future research in this area might focus on educational factors that produce good evidence-based reasoners. One possible direction for future research would be to provide training on predictive factors that can be influenced by experience and training. As these predictive factors are improved, evidence-based reasoning should improve as well. Conversely, evidence-based reasoning could be taught or trained directly. In my studies, history students appear to be ahead of the crowd in terms of evidence-based reasoning; however, further research is needed to better understand these educational differences and what contributes to them. If it is the case that people in some fields are trained to be better evidence-based reasoners than others, is there something to be learned from the way that these people are being trained?

On the other hand it could be that, as with the history vs. psychology students, people are more critical of evidence that is not familiar to them just as they are more critical of evidence with which they disagree (Lord, Ross & Lepper, 1979). This did seem to be the case in study 1, where undergraduates provided more evaluations of unfamiliar (cognitive psychology) vignettes as opposed to familiar (general knowledge) vignettes. This distinction between reasoning in familiar and unfamiliar contexts could be used as a teaching tool. Previous research has shown that contrasting cases can be a powerful way to get people thinking carefully and actively (Schwartz & Bransford, 1998). Perhaps one way to train people to be more critical would be to present them with evidence from a familiar and unfamiliar discipline, and to point out differences in the way the two types of evidence are interpreted.

Another potential area for fruitful research would be to examine how context affects the ways that people interpret data. In the current studies, the rate at which individuals' critically evaluate data, even when explicit and major flaws are built into studies, is arguably small (about 20% of the flaws were noted when individuals were asked to discuss research findings, and 27% when asked to critically evaluate research findings). Future research might consider how various characteristics of research vignettes influence the likelihood that individuals critically evaluate versus simply explain/relate data to prior knowledge. One factor that might influence how critical individuals are may be the existence of graphs depicting the data. I was unable to compare data from study 1 (where graphs accompanied vignettes) and study 2 (where vignettes were presented alone) because of many other differences between the two studies. However, one study by Fagerlin, Ubel & Wang (2005) in a medical decision making context found that individuals were less influenced by personal anecdotes and more influenced by statistical information when data were presented in a graph compared to in-text. Similarly, individuals may be more focused on the actual data and less likely to activate their prior experiences if study descriptions include visual displays of data. Such visual displays may also highlight certain flaws by making them more visually salient (e.g., small effect sizes) or by emphasizing the key variables in the study as opposed to the factors considered in a broader interpretation of data. While graphs may help individuals focus on the data and perhaps lead them to be more critical of data, it is plausible that personal anecdotes included in a media story about a research findings may have the impact of interfering with individuals' focus on the data and, therefore, making them less critical.

Also, the credibility and trustworthiness of the source of the research findings also has a likely influence on the critical evaluation of data (Petty, Priester, & Brinol, 2002). Prior research has found that the trustworthiness of sources influence how likely individuals are to be persuaded by messages depending on the source (Andreoli & Worchel, 1978). Furthermore, research has shown that individuals are more likely to elaborate upon and critically examine ideas when they come from a less trustworthy source than a more trustworthy source (Priester& Petty, 1995). It follow from this research that individuals are likely to be more critical of scientific data when the source is less trustworthy. Identification of such factors that influence how critically individuals read research findings might help us to gain a better understanding of situations in which people are likely to be more or less critical and could have application in a wide range of contexts such as science journalism, textbook writing and classrooms.

People are consistently placed in situations in which they need to use data to make decisions. Many believe that, in order for people to fully participate as members of their communities and societies, they must be able to use data to think carefully and actively about a wide range of issues. Additionally, increased economic competition associated with globalization means that people are being exposed to more complex, technologically driven environments. Researchers, educators, educational policy-makers, and employers agree that students need to learn to "think well" now more than ever before. I hope that my work makes a significant contribution to the body of knowledge on how people can and do think critically in their everyday lives and what underlies this skill.

Table 1

Study 1 Descriptive Statistics

Vignette Tasks	# of Quantitative Descriptions (SD)	# of Explanations (SD)	# of Evaluations (SD)
Undergrads	5.33 (4.45)	5.22 (4.94)	3.63 (4.48)
Soc Psych Grads	12.43 (5.26)	2.43 (2.63)	5.00 (5.35)
History Grads	4.75 (5.09)	1.62 (2.07)	7.75 (6.02)
ME Grads	7.00 (5.04)	1.91 (1.51)	9.00 (4.27)
Cog Psych Grads	9.22 (5.49)	2.56 (4.00)	4.89 (2.80)
Total	6.54 (5.17)	3.88 (4.36)	5.02 (4.87)

Table 2

Pilot Study Descriptive Statistics

Task	Mean (SD)	
AOT Questionnaire	149.55 (19.37)	
NFC Questionnaire	66.45 (10.03)	
Vignette Task	Discuss studies	Critically evaluate
# of Evaluations	3.18 (2.42)	4.06 (2.55)
# of Flaws Recognized	1.51 (1.91)	2.06 (1.82)
# of Explanations	5.56 (2.54)	6.03 (2.04)

Table 3

Participant Frequencies for Study 2

	Engineering	History	Psychology	Undeclared	Other	Total
Freshman	1	0	4	47	26	78
Sophomore	10	4	8	28	28	78
Junior	12	18	12	2	7	51
Senior	13	31	9	0	8	61
Total	36	53	33	77	69	268

Table 4

ANOVA Results for Lab vs. Non-lab Participants

Vignette Task	Degree of Freedom	F Value	Sig (*p*)	Partial Eta Squared
Discuss Studies: Evaluations	1, 29	1.25	.265	.005
Discuss Studies: Flaws Recognized	1, 29	.045	.831	.000
Discuss Studies: Explanations	1, 29	.538	.464	.002
Critically Evaluate: Evaluations	1, 29	1.44	.231	.005
Critically Evaluate: Flaws Recognized	1, 29	.290	.591	.001
Critically Evaluate: Explanations	1, 29	.046	.831	.000

Table 5

Study 2 Descriptive Statistics

	Mean	Standard Deviation	Range
NFC	62.08	10.48	53.00
AOT	150.70	19.63	104.00
EBQ	91.59	9.59	92.00
Percentile SAT/ACT	91.53	9.35	59.00
Media Interest	4.87	1.11	5.00
Science Media Interest	4.78	1.66	6.00
Critical of Studies	4.14	1.68	6.00
Fam. w/ Corr/ Caus	4.18	0.99	4.00
Fam w/ Oper.	3.76	1.18	4.00
Fam w/ Effect Size	3.69	1.09	4.00
Fam w/ Samp Bias	4.33	0.88	4.00
Discuss Studies # Evals	3.45	2.35	8.00
DS # Flaws Recognized	1.60	1.47	7.00
DS # Expls	2.88	2.01	8.00
Crit Evaluate # Evals	5.18	2.16	8.00
CE # Flaws Recognized	2.13	1.48	7.00
DS # Flaws Recognized	3.24	2.37	8.00
# of Stats Courses	0.60	0.84	6.00

Table 6

Vignette Descriptive Statistics by Year in College

	1st instructions: Discuss the studies			2nd instructions: Critically evaluate		
	# of evaluations (SD)	# of recognitions (SD)	# of explanations (SD)	# of evaluations (SD)	# of recognitions (SD)	# of explanations (SD)
Freshman	2.56 (2.07)	1.19 (1.18)	3.04 (2.13)	4.87 (2.27)	1.77 (1.34)	3.19 (2.27)
Sophomore	3.13 (2.23)	1.35 (1.30)	3.12 (2.08)	4.69 (2.16)	2.03 (1.44)	3.33 (2.51)
Junior	3.86 (2.27)	1.63 (1.34)	2.76 (1.76)	5.27 (2.15)	2.02 (1.56)	3.41 (2.36)
Senior	4.66 (2.36)	2.41 (1.79)	2.46 (1.95)	6.10 (1.76)	2.82 (1.45)	3.03 (2.37)

Table 7

Vignette Descriptive Statistics by Major

	1st instructions: Discuss the studies			2nd instructions: Critically evaluate		
	# of evaluations (SD)	# of recognitions (SD)	# of explanations (SD)	# of evaluations (SD)	# of recognitions (SD)	# of explanations (SD)
Engineering	3.97 (2.43)	1.81 (1.74)	3.17 (1.96)	5.33 (1.84)	2.06 (1.58)	3.17 (2.37)
History	4.75 (2.19)	2.17 (1.48)	2.70 (1.98)	5.89 (1.90)	2.66 (1.47)	3.17 (2.44)
Psychology	2.94 (2.30)	1.48 (1.60)	2.91 (2.24)	5.00 (2.18)	2.27 (1.33)	3.61 (2.32)
Undeclared	2.70 (2.05)	1.21 (1.20)	2.90 (2.07)	4.90 (2.31)	1.88 (1.48)	3.27 (2.22)
Other	3.26 (2.34)	1.54 (1.42)	2.83 (1.93)	4.94 (2.26)	1.97 (1.44)	3.12 (2.55)

Table 8

Questionnaire Descriptive Statistics by Year in College

	NFC Mean (SD)	AOT Mean (SD)	EBQ Mean (SD)	Percentile SAT/ACT Mean (SD)
Freshman	58.56 (9.02)	145.53 (20.60)	89.96 (9.51)	90.24 (9.18)
Sophomore	62.82 (11.22)	150.21 (18.62)	92.14 (8.70)	92.19 (7.98)
Junior	63.20 (11.08)	152.22 (19.62)	91.63 (10.52)	90.12 (12.45)
Senior	64.69 (9.77)	156.67 (18.17)	92.93 (9.91)	93.54 (7.89)

Table 9

Disposition, Belief and Ability Questionnaire Descriptive Statistics by Major

	NFC Mean (SD)	AOT Mean (SD)	EBQ Mean (SD)	Percentile SAT/ACT Mean (SD)
Engineering	65.25 (9.77)	153.67 (20.48)	90.44 (11.78)	94.41 (4.74)
History	64.96 (11.07)	159.49 (20.24)	95.09 (7.95)	92.74 (8.96)
Psychology	61.58 (9.25)	146.55 (16.20)	90.24 (8.74)	92.12 (7.03)
Undeclared	58.90 (10.19)	148.36 (20.14)	90.75 (7.53)	90.36 (10.89)
Other	62 (10.42)	146.99 (17.69)	91.07 (11.39)	90.22 (10.14)

Table 10

Interest Item Descriptive Statistics by Year in College

	Media Interest Mean (SD)	Science Media Interest Mean (SD)	Critical of Studies Mean (SD)
Freshman	4.65 (1.15)	4.63 (1.67)	3.81 (1.74)
Sophomore	4.79 (1.10)	4.94 (1.56)	4.09 (1.59)
Junior	5.08 (.96)	5.00 (1.64)	4.48 (1.52)
Senior	5.07 (1.14)	4.59 (1.78)	4.33 (1.77)

Table 11

Interest Item Descriptive Statistics by Major

	Media Interest Mean (SD)	Science Media Interest Mean (SD)	Critical of Studies Mean (SD)
Engineering	5.17 (.941)	5.89 (1.26)	4.89 (1.45)
History	5.28 (.885)	3.92 (1.69)	3.91 (1.70)
Psychology	4.52 (1.35)	4.91 (1.63)	4.45 (1.72)
Undeclared	4.75 (1.11)	4.70 (1.61)	3.94 (1.69)
Other	4.70 (1.10)	4.88 (1.56)	4.00 (1.64)

Table 12

Familiarity with Statistics/Methodology Item Descriptive Statistics by Year in College

	Correlation/ Causation Mean (SD)	Operationalization Mean (SD)	Effect Size Mean (SD)	Sampling Bias Mean (SD)
Freshman	4.04 (1.04)	3.40 (1.23)	3.32 (1.21)	4.18 (.922)
Sophomore	4.03 (1.04)	3.74 (1.19)	3.62 (.996)	4.24 (.983)
Junior	4.06 (1.09)	3.90 (1.14)	3.90 (.922)	4.41 (.753)
Senior	4.64 (.578)	4.15 (1.01)	4.10 (.995)	4.56 (.742)

Table 13

Familiarity with Statistics/Methodology Item Descriptive Statistics by Major

	Correlation/ Causation Mean (SD)	Operationalization Mean (SD)	Effect Size Mean (SD)	Sampling Bias Mean (SD)
Engineering	4.31 (1.04)	4.14 (1.02)	4.11 (.95)	4.50 (.66)
History	4.28 (.84)	3.92 (1.11)	3.81 (.94)	4.38 (.81)
Psychology	4.42 (.87)	3.88 (1.27)	4.27 (.84)	4.73 (.52)
Undeclared	4.04 (1.03)	3.49 (1.23)	3.40 (1.13)	4.19 (1.00)
Other	4.06 (1.07)	3.70 (1.17)	3.43 (1.14)	4.16 (.96)

Table 14

Percentage of People who Recognized Flaw in Each Vignette by Year in College and Major

	Fresh	Soph	Jun	Sen	Eng	His	Psych	Und	Oth
Obesity	3.8	11.5	3.9	23	8.3	17.0	9.1	5.2	13
Creativity	7.7	14.1	17.6	21.3	25.0	18.9	9.1	10.4	13
Abuse	10.3	12.8	25.5	34.4	19.4	34	12.1	9.1	23.2
Runners	10.3	10.3	5.9	19.7	8.3	5.7	24.2	10.4	13
IQ	6.4	6.4	15.7	19.7	19.4	13.2	15.2	6.5	8.7
Video Games	29.5	26.9	31.4	49.2	36.1	49.1	27.3	27.3	30.4
Day Care	1.3	5.1	7.8	3.3	8.3	3.8	6.1	1.3	4.3
Politics	50	47.4	54.9	70.5	55.6	75.5	45.5	50.6	47.8

Table 15

Correlations with Non-Vignette Data

	1	2	3	4	5	6	7	8	9	10	11	12
1. EBQ	1	.323**	.320**	.134*	-.009	.020	.089	.103	.171	.038	.122*	.021
2. NFC	.323**	1	.400**	.099	.261**	.081	.215**	.215**	.270**	.213**	.095	.132*
3. AOT	.320**	.400**	1	.018	.185**	.043	.161**	.182**	.332**	.182**	.098	.154*
4. # Stats Classes	.134*	.099	.018	1	.103	-.028	-.047	.091	.196**	.138*	.291**	.157*
5. Percentile SAT/ACT	-.009	.261**	.185**	.103	1	.107	.138*	.130*	.244*	.172**	.051	.209**
6. Media Interest	.020	.081	.043	-.028	.107	1	.146*	.062	.072	.114	.035	.110
7. Science Media Interest	.089	.215**	.161**	-.047	.138*	.146*	1	.597**	.183**	.153*	.089	.106
8. Critical of Studies	.103	.215**	.182**	.091	.130*	.062	.597**	1	.136*	.144*	.182**	.113
9. Fam. w/ Corr/ Caus	.171**	.270**	.332**	.196**	.244**	.072	.183**	.136*	1	.528**	.321**	.435**
10. Fam w/ Oper.	.038	.213**	.182**	.138*	.172**	.114	.153*	.144*	.528**	1	.528**	.435**
11. Fam w/ Effect Size	.122*	.095	.098	.291**	.051	.035	.089	.182**	.321**	.528**	1	.305**
12. Fam w/ Samp Bias	.021	.132*	.154*	.157*	.209**	.110	.106	.113	.435**	.435**	.305**	1

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 16

Correlations with Vignette Data

Out of 8 Vignettes	Discuss Stud. # Evals	DS # Flaws Recognized	DS # Expls	Crit Evaluate # Evals	CE # Flaws Recognized	CE # Expls
EBQ	.178**	.112	.093	.079	.070	.097
NFC	.269**	.238**	-.012	.262**	.238**	-.011
AOT	.209**	.271**	.038	.170**	.314**	.077
# Stats Classes	.133*	.126*	-.102	.172**	.141*	.010
Percentile SAT/ACT	.190**	.181**	-.081	.157*	.207**	.011
Media Interest	.053	.108	-.002	.021	.061	.046
Science Media Interest	.067	.062	.038	.045	.077	.085
Critical of Studies	.133*	.123*	-.059	.082	.112	.022
Fam. w/ Corr/ Caus	.246**	.256**	-.049	.198**	.252**	.103
Fam w/ Oper.	.183**	.148*	-.044	.169**	.172**	-.020
Fam w/ Effect Size	.078	.073	-.009	.066	.078	-.076
Fam w/ Samp Bias	.196**	.206	-.002	.119	.176**	.036

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 17

Non-significant Interaction Effects: x Amount of Education x Condition (Discuss Study vs. Critically Evaluate) and Major x Condition (Discuss Study vs. Critically Evaluate)

Interaction effect	Degree of Freedom	F Value	Sig (*p)*	Partial Eta Squared
# of Evaluations: Effect of Amount of Education x Condition	3, 264	2.04	.109	.023
# of Recognitions: Effect of Amount of Education x Condition	3, 264	.633	.594	.007
# of Explanations: Effects of Amount of Education x Condition	3, 264	.548	.650	.006
# of Evaluations: Effect of Major x Condition	4, 263	1.76	.137	.026
# of Recognitions: Effect of Major x Condition	4, 263	.930	.447	.014
# of Explanations: Effects of Major x Condition	4, 263	.325	.861	.005

Figure 1. Study 1: Descriptions of Data (out of 16)

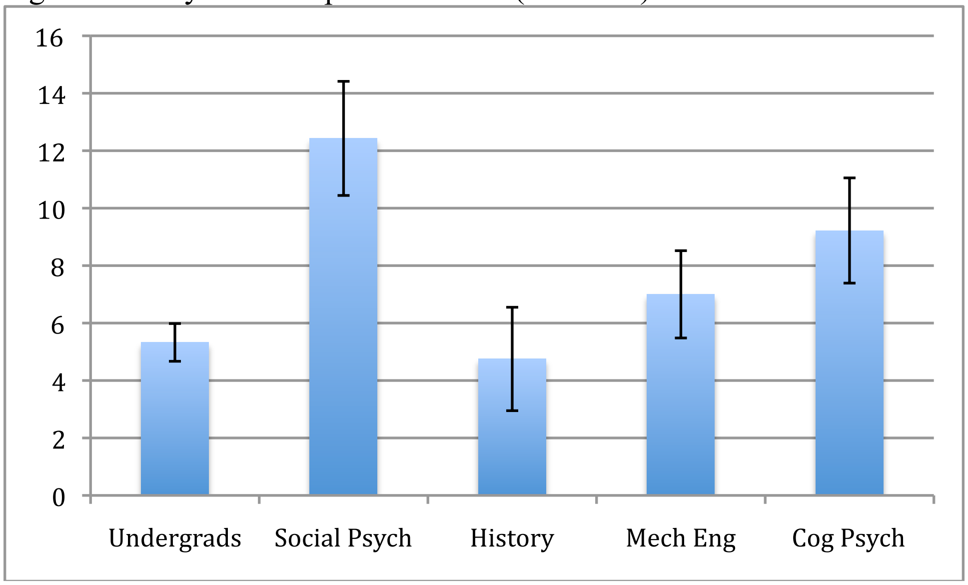


Figure 2. Study 1: Evaluations of Data (out of 16)

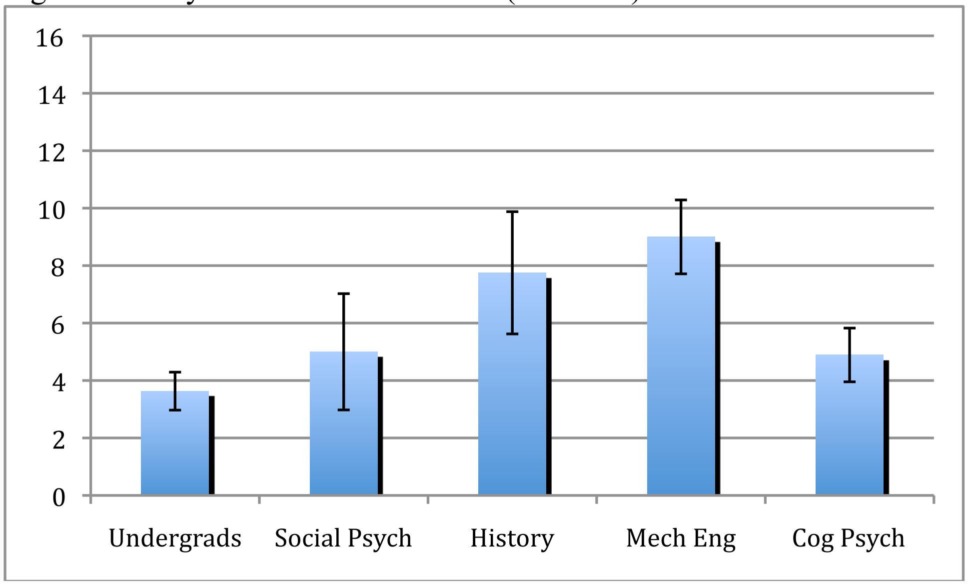


Figure 3. Study 1: Explanations of Data (out of 16)

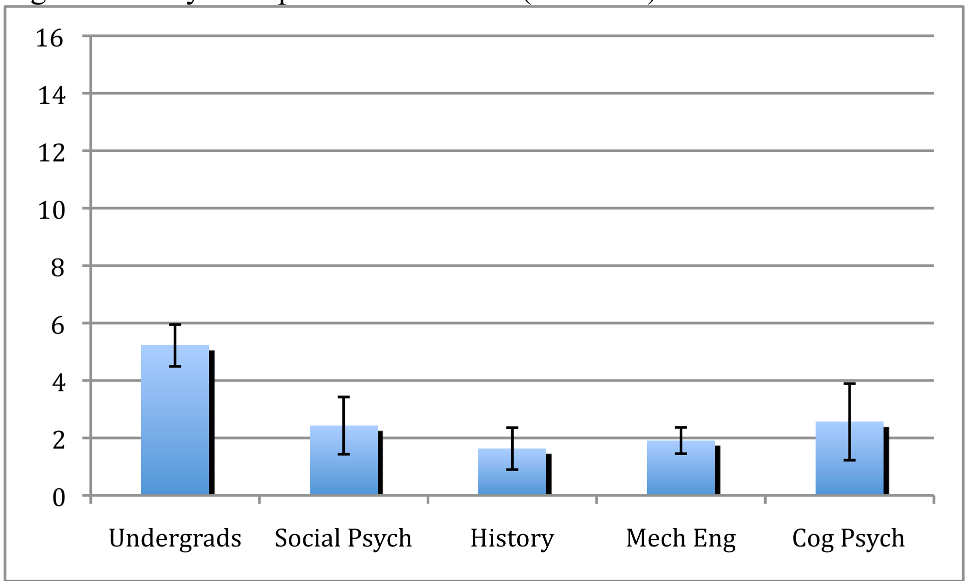


Figure 4. Study 1: Descriptions of Data by Type of Graph

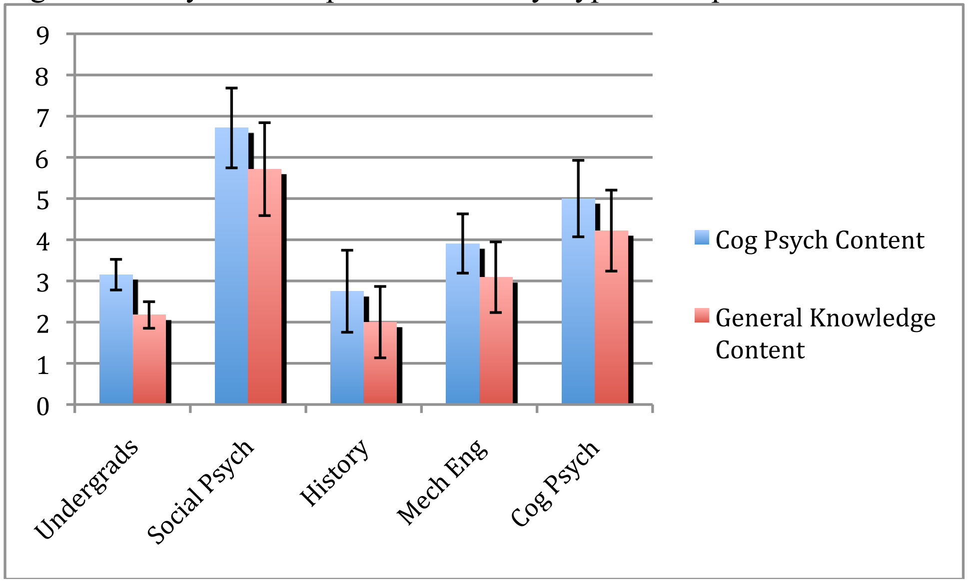


Figure 5. Study 1: Evaluations of Data by Type of Graph

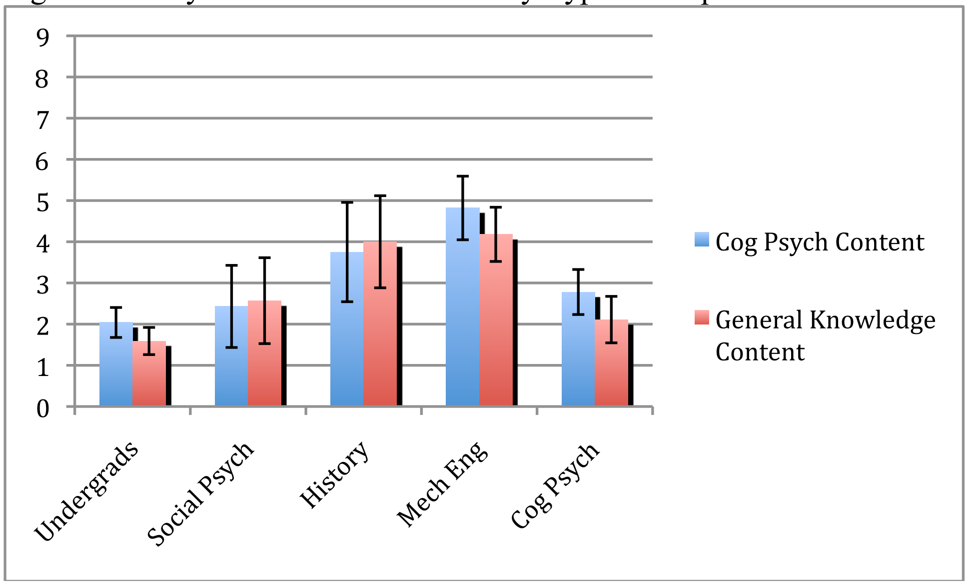


Figure 6. Study 1: Explanations of Data by Type of Graph

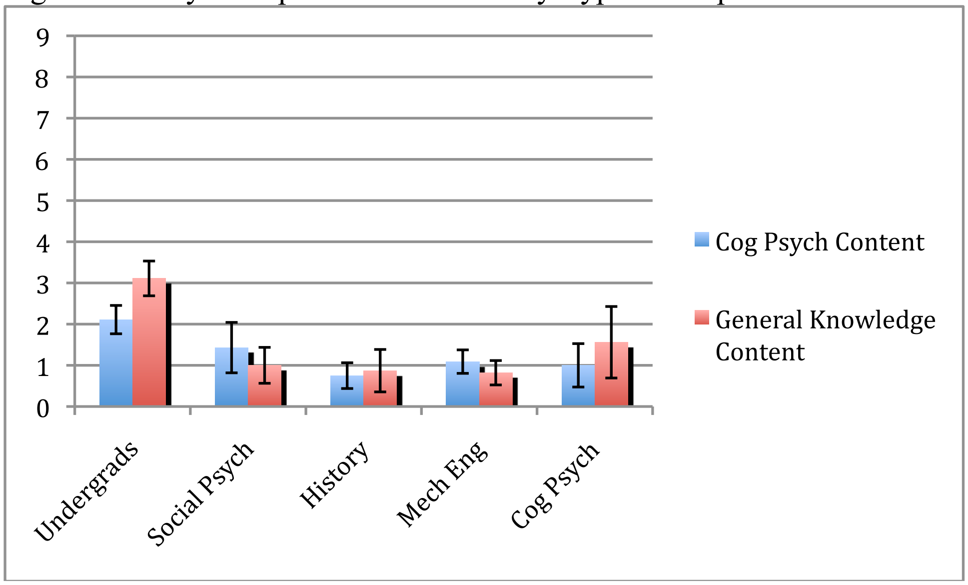


Figure 7. Study 2: Vignettes by Type of Instruction

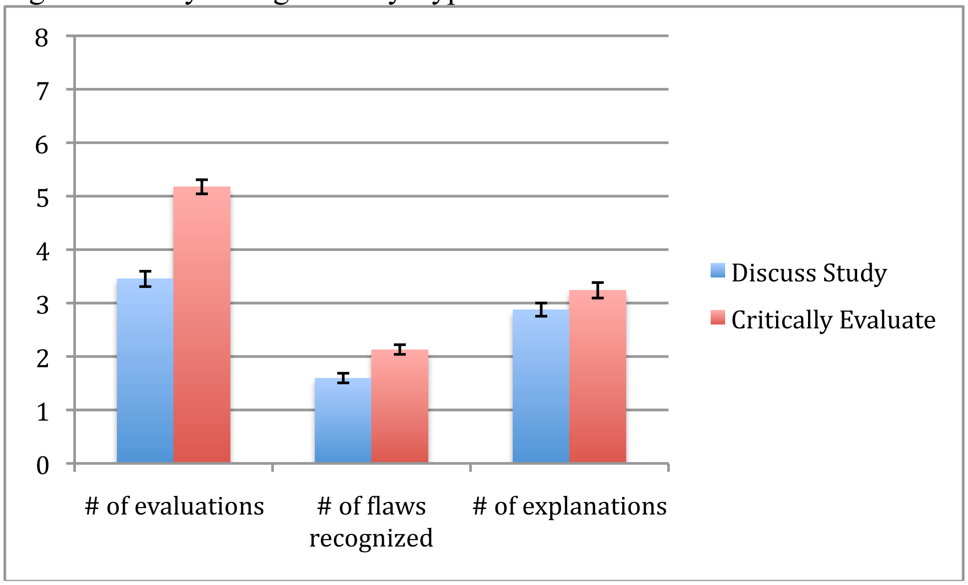


Figure 8. Study 2: Vignettes by Type of Flaw

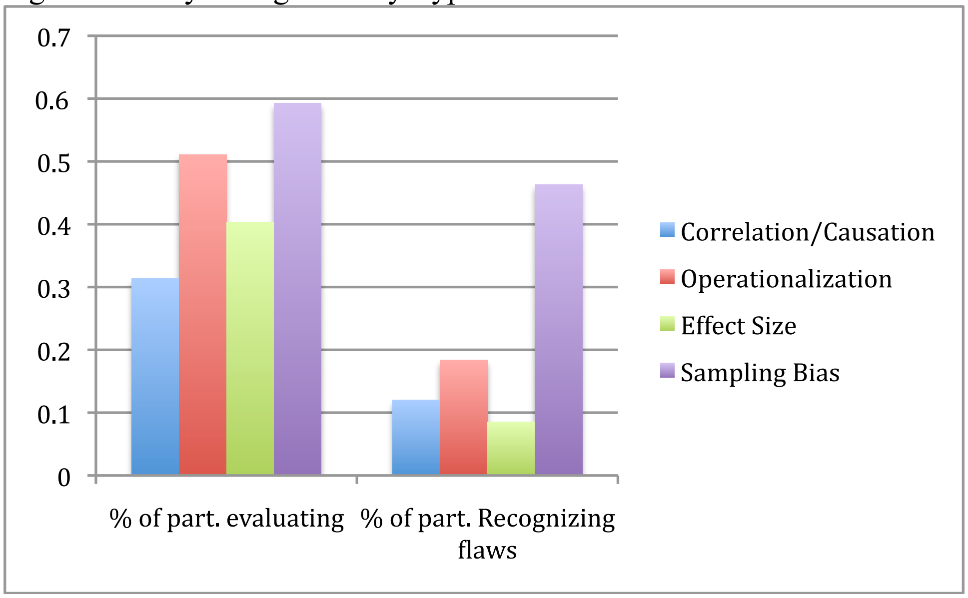


Figure 9. Study 2: Vignette Interpretation by Year in College ("Discuss Studies" Condition)

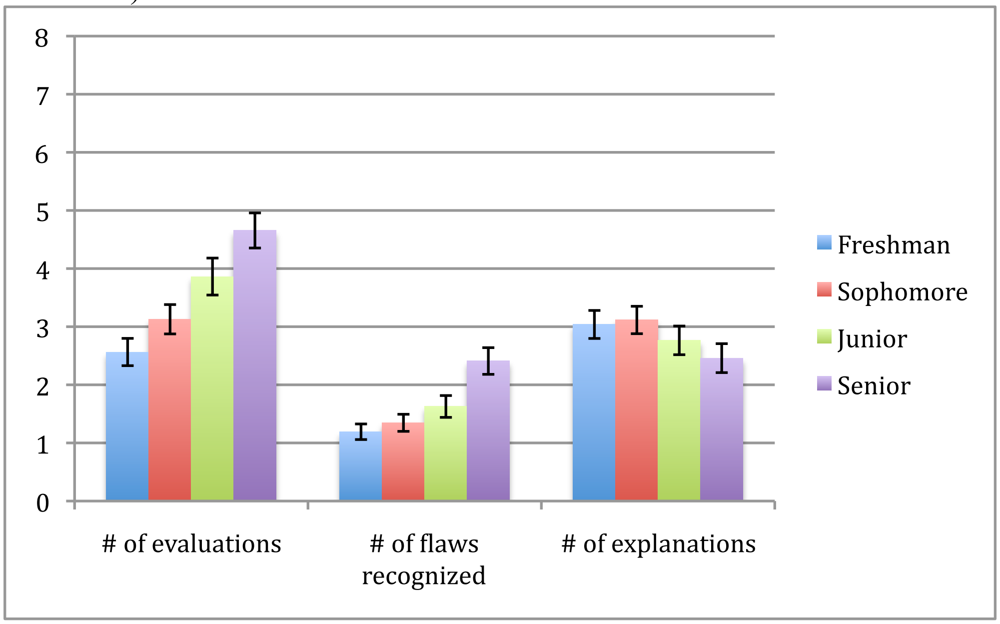
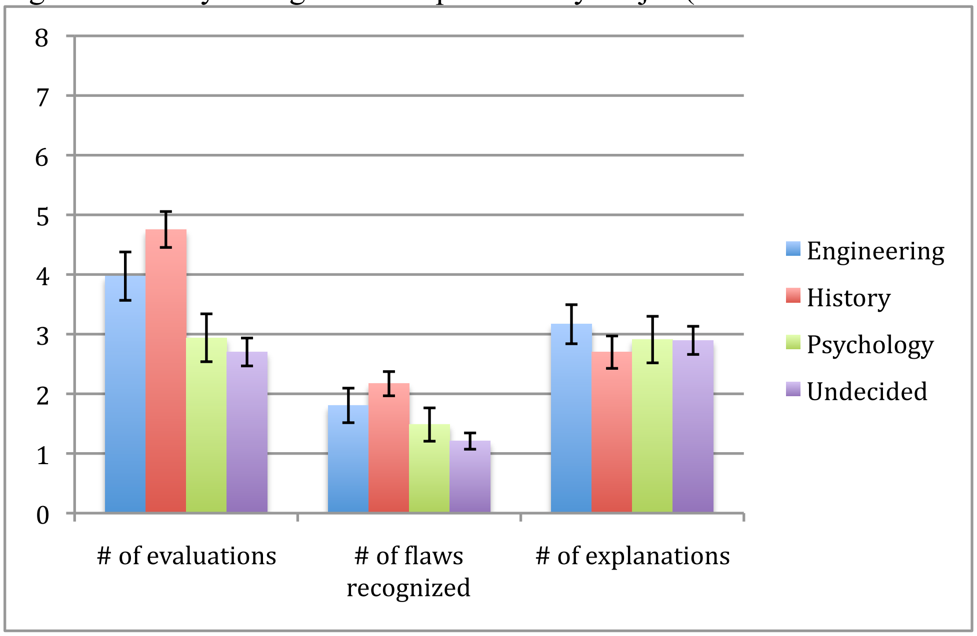


Figure 10. Study 2: Vignette Interpretation by Major ("Discuss Studies" Condition)

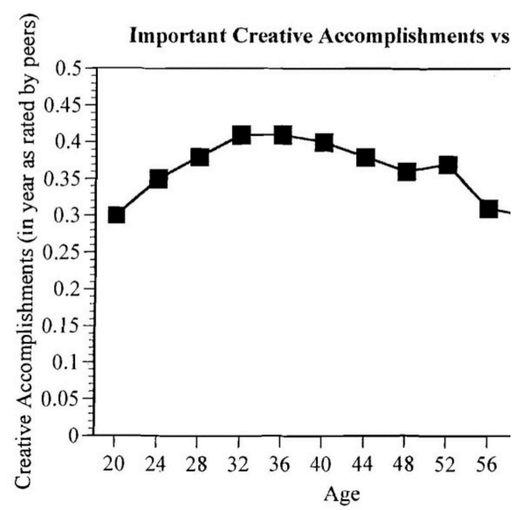


Appendices

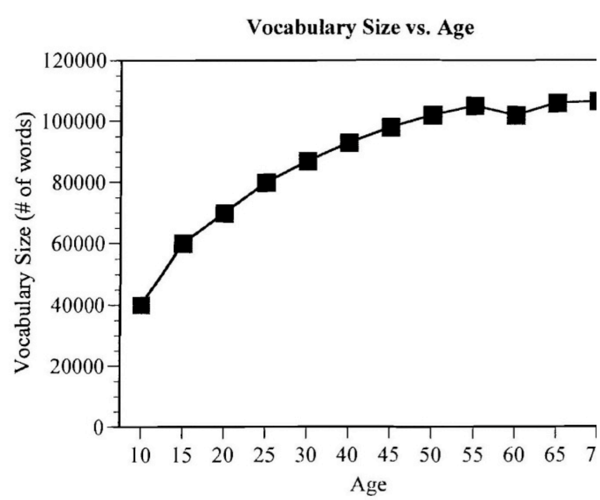
Appendix A

Study 1 Vignettes

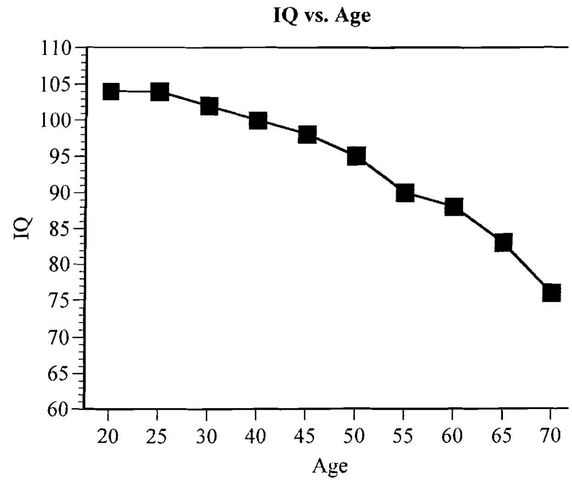
A psychologist interested in how creativity in scientific reasoning changes with conducted a study. He asked colleagues of hundreds of randomly chosen scientists from across the United States to rank the mean number of HIGHLY SIGNIFICANT creative, scientific accomplishments that their colleagues they had made over the past year. He calculated the average number of creative accomplishments as a function of age. On the right are the results of his study. Please discuss this study.



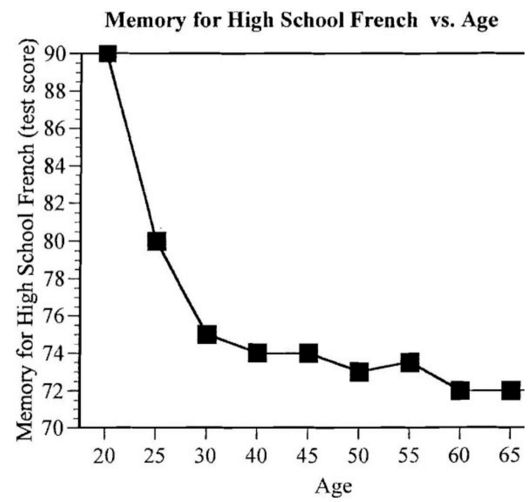
A developmental psycholinguist was interested in the size of vocabulary as people get older. She used a classic procedure for estimating vocabulary size in which one counts the number of words in a dictionary, and asks people to define and use a random sample of those words. They then multiply the percentage of words they knew by the total number of words possible to estimate total vocabulary. She conducted her study on a large number of people aged 10-70. On the right are her results. Please discuss the study.



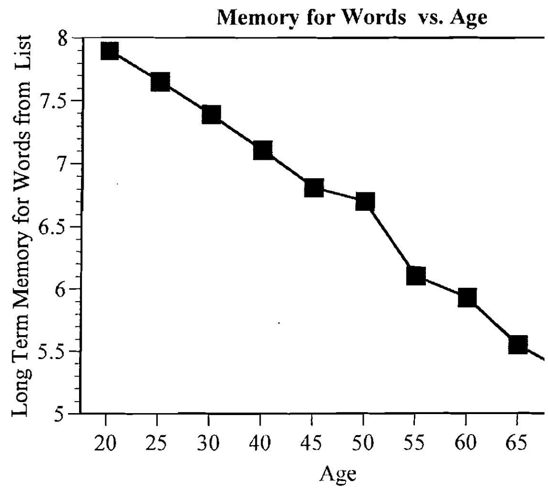
A cognitive aging researcher was interested in how IQ changes as we get older. She gave a large number of people of different ages a standard battery of IQ tests and calculated their IQ's. On the right are her results. Please discuss the study.



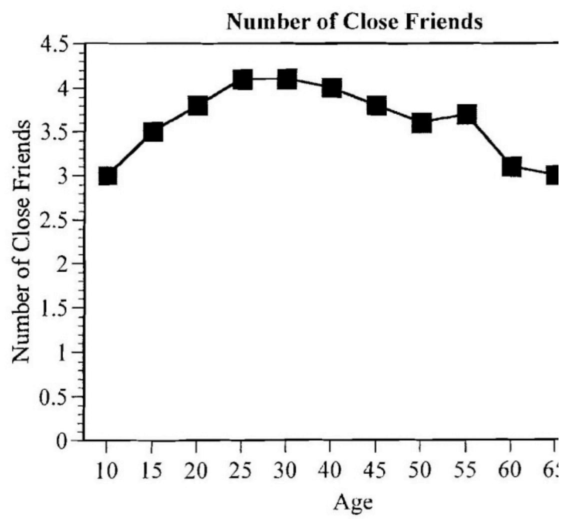
A famous study examined people's memory for hte information they learned in their high school French courses as they got older. A psychologist with an interest in very long-term memory was interested in replicating the results of that study. She conducted a longitudinal study over fifty years in which she gave people tests of high school French every five years. On the right are her results. Please discuss the study.



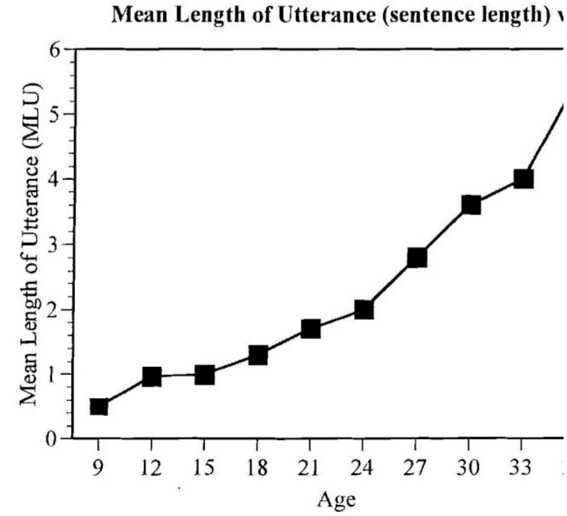
An aging resarcher was interested in how their long-term memory for lists changes as people get older. She gave people of different ages a list of 20 words and asked them to recall as many words as they could one week later. On the right are the results of her study. Please disucss the study.



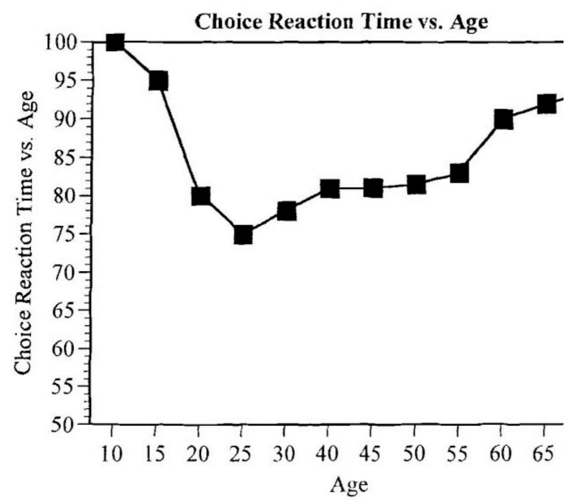
A researcher at the Institute for Social Research was interested in whether people's social support networks changed in size as they got older. He asked people of different ages how many friends they had that they considered "very close." On the right are is results. Please discuss the study.



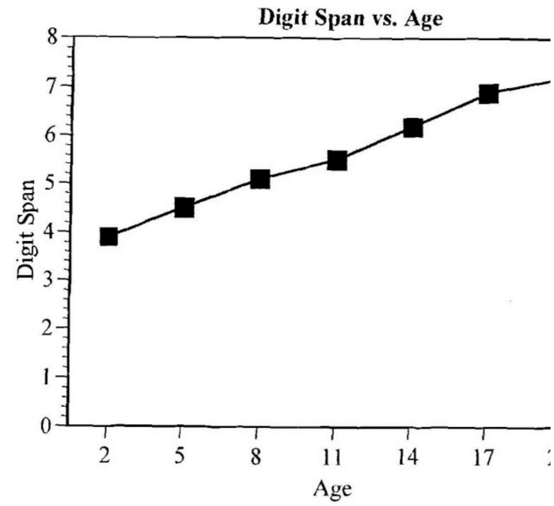
A developmental psycholinguist was interested in the mean length of utterance of children as they get older (that is, the number of words they say in a single "sentence." Children start of saying one word at a time, then two words in a phrase, etc. He videotaped a large number of children for several hours and then transcribed their speech (or lack of speech) and calculated the average length of utterances at each age range. On the right are his results. Please discuss the study.



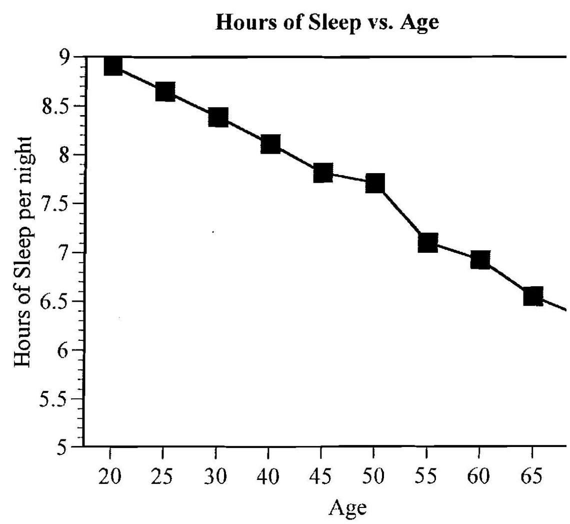
A cognitive psychology class decided to find out how people's reaction time speeds change with age-- that is, they were interested in how quickly can on respond to incoming information? A large number of participants aged 10-70 were recruited to perform a task on a computer in which they were to respond by pressing a button on the left side if a circle appeared on the screen, and a button on the right side if a square appeared on the screen. The average time (in milliseconds) it took for each age group is shown on the right. Please discuss this study.



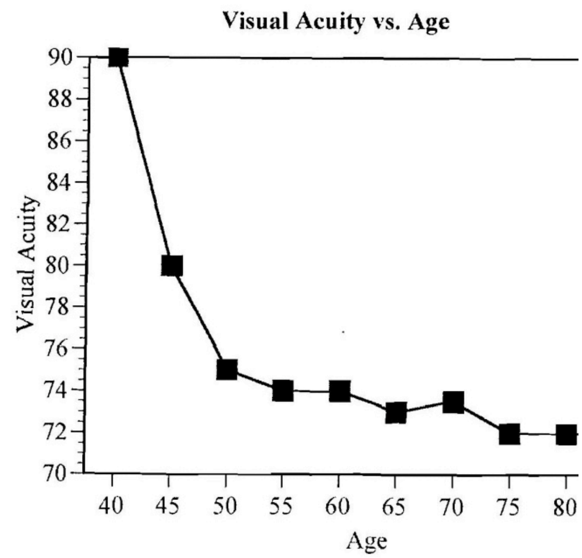
An developmental psychologist was interested in how people's short-term memory changes as they get older. He gave children and young adults lists of numbers that they were asked to recall immediately following presentation. The most numbers that a person can remember is referred to as their "digit span." The researcher calculated the average digit span for each age group. On the right are results of his study. Please disucss the study.



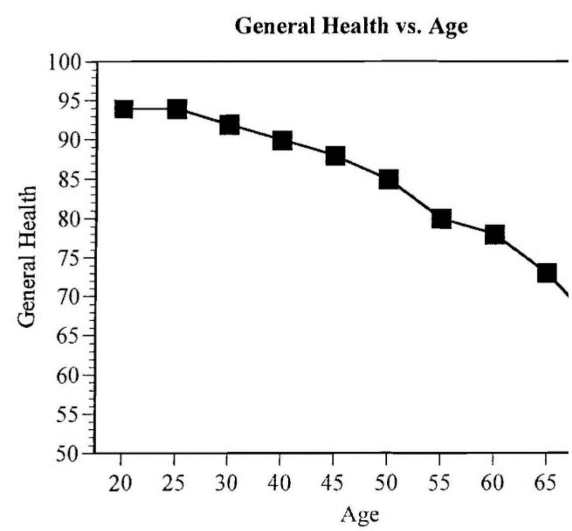
A scientist at the sleep institute was interested in how many hours people slept a night as a function of age, so he surveyed a large number of people of different ages. On the right are his results. Please discuss the study.



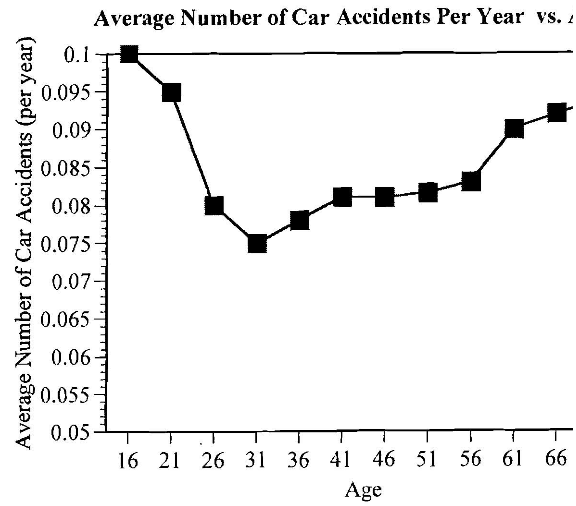
A research ophthalmologist at the Kellogg Eye Center conducted a study on the change in visual acuity with age. He examined the visual acuity of a total of 500 patients (50 at each of the following ages:40, 45, 50, 55, 60, 65, 70, 75, 08, 85). He calculated an overall score of their visual acuity that ranged from 1-100. On the right are the results of his study, in which his visual acuity score is plotted as a function of age. Please discuss this study.



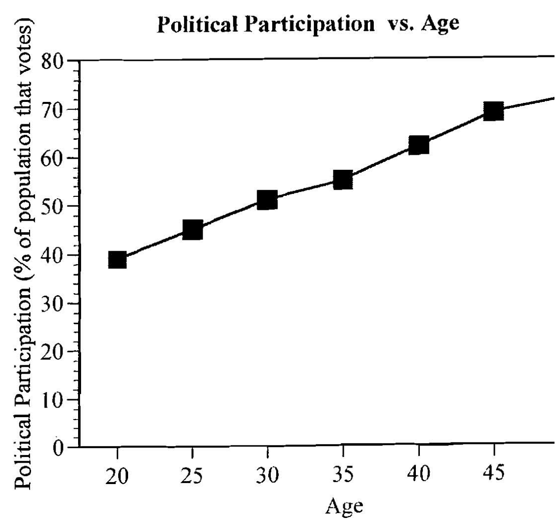
A physician was interested in the pattern of general health changes with time. She surveyed her patients about a large number of possible health problems, and also health factors (weight, blood pressure, etc) and gave them an overall "general health" score based on the number of "good" and "bad" factors involved. She plotted their general health score as a function of age on the right. Please discuss the study.



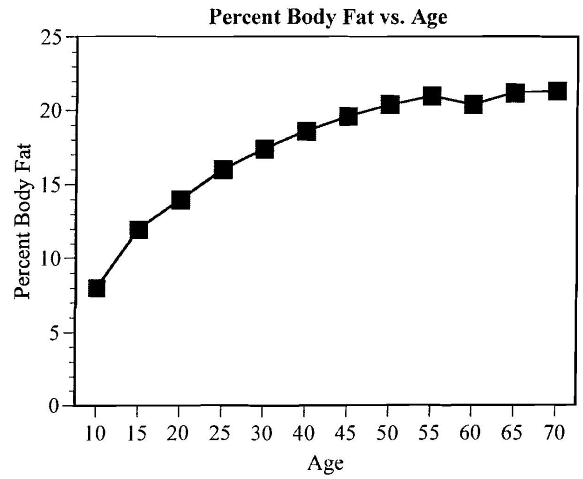
The State Department of Motor Vehicles was interested in computing car accident statistics for the entire state of Michigan so they collected reports from police departments across the state for the ten years and computed the average number of car accidents, per licensed driver of a particular age, per year. The results of their study are plotted on the right. Please discuss the study.



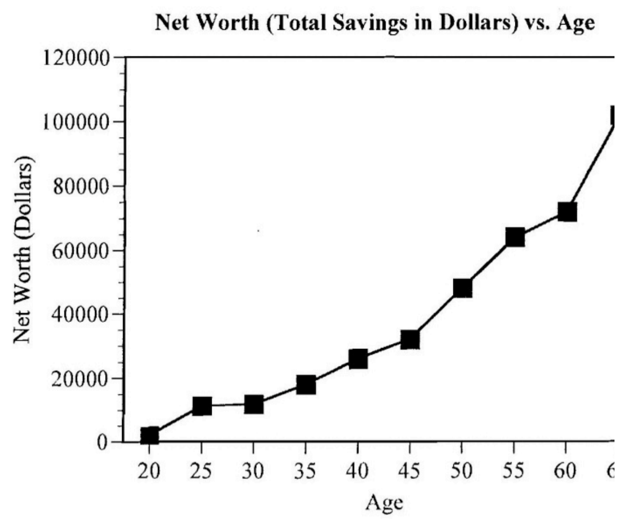
A journalist for CBS news conducted a large-scale poll immediately following the Presidential election to find out how political participation changed with age. He calculated the percent of people at each age that voted in the election. The results are plotted on the right. Please discuss the study.



A nutritionist was interested in how body fat changes with age, and she did a study in which she measured hte percent of body fat of a large number of people. Her results are plotted on the right. Please discuss the study.



An economist was intersted how people's total net worth (savings in dollars) changed with age, so she conducted a survey in which she asked people how much money they had (including equity in home, retirement savings, etc). Her results are plotted on the right. Please discuss this study.



Appendix B

Pilot Study Vignettes

Misinterpretation of correlational data (i.e. correlation/causation confusion)

A study of 77 children aged 3 to 5 found that those with the most body fat had the most “controlling” mothers when it came to the amount of food eaten. “The more control the mother reported using over her child’s eating, the less self-regulation the child displayed,” Dr. Johnson and her co-author said.

A recent study followed the health of 481 members of a runners’ club and 330 nonrunners living in the same community. All participants were 50-72 when the study began. They found that long-distance running was associated with good physical function in the later years, compared with more sedentary life styles. They said the findings underscored the importance of promoting “regular lifetime physical exercise to improve the quality of life of the growing older population.”

Unclear operationalization of variables

A researcher interested in how creativity in scientific reasoning changes with age conducted a study. He asked colleagues of hundreds of randomly chosen scientists from across the United States to rank the mean number of creative scientific accomplishments that their colleagues had made over the past year. He calculated with average number of creative accomplishments as a function of age, and found that number of creative accomplishments increased from age 20-35 and then began to decline.

One in four adolescents said they were abused within the past year, according to a new survey. The telephone survey of 2,000 children ages 10 to 16, suggests, “We’re not doing a very good job of counting and tracking the problem,” said David Finkelhor, a sociologist at the University of New Hampshire and co-author of the study.

Over-interpretation of small effect sizes

The eldest children in families tend to develop higher I.Q.'s than their siblings, researchers are reporting today, in a large study (two papers) that could settle more than a half-century of scientific debate about the relationship between I.Q. and birth order. The average difference in I.Q. was three points higher in the eldest child than in the closest sibling. "I consider these two papers the most important publications to come out in this field in 70 years; it's a dream come true," said Frank J. Sulloway, a psychologist at the Institute of Personality and Social Research at the University of California, Berkeley.

A study of American child care has found that keeping a preschooler in a day care center for a year or more increased the likelihood that the child would become disruptive in class -- and that the effect persisted through the sixth grade. Every year spent in such centers for at least 10 hours per week was associated with a 1 percent higher score on a standardized assessment of problem behaviors completed by teachers, said Dr. Margaret Burchinal, a co-author of the study and a psychologist at the University of North Carolina. With more than two million American preschoolers attending day care, the increased disruptiveness very likely contributes to the load on teachers who must manage large classrooms, the authors argue.

Biased samples

A recent study shows that Americans are more interested in politics than was previously thought. Researchers approached people at events such as town meetings and city council meetings and surveyed them about their interest in political and their voting behavior. Seventy percent of those surveyed reported that they were planning to vote in upcoming local and national elections. More than half said that they regularly read articles in the newspaper about political issues.

A study conducted by Market Research International shows that educational video games are just as popular as non-educational video games. The researchers conducted their study at a local university with children of university professors. They allowed the children to play a variety of educational and non-educational video games. They then measured the amount of time the children spent playing the two types of games. Interestingly, the amount of time that children spent playing educational video games was not significantly different from the amount of time they spent playing non-educational video games.

Appendix C

Study 2 Vignettes

Misinterpretation of correlational data (i.e. correlation/causation confusion)

As part of a recent study, researchers measured children's body fat and surveyed mothers about the amount of control they exert over their children's eating. The results of this study, conducted with 77 children aged 3 to 5, found that those with the most body fat had the most "controlling" mothers when it came to the amount of food eaten. This shows that when mothers exert more control over their children's eating, the children display less self-control, researchers said.

A recent study followed the health of 481 members of a runners' club and 330 nonrunners living in the same community. All participants were 50-72 when the study began. Researchers found that long-distance runners had better physical function in the later years compared with those with more sedentary life styles. They said the findings underscored the importance of promoting regular lifetime physical exercise to improve the health of the growing older population.

Unclear operationalization of variables

A researcher interested in how creativity in scientific reasoning changes with age conducted a study. He asked colleagues of hundreds of randomly chosen scientists from across the United States to rank the mean number of creative scientific accomplishments that their colleagues had made over the past year. He calculated the average number of creative accomplishments as a function of age, and found that number of creative accomplishments increased from age 20-35 and then began to decline.

One in four adolescents said they were abused within the past year, according to a new survey. As part of the telephone survey of 2,000 children ages 10 to 16, participants were asked to report whether they had been abused in the past year. The result suggests that we have not been doing a very good job of counting and tracking the problem, the researchers said.

Over-interpretation of small effect sizes

The eldest children in families tend to develop higher I.Q.s than their siblings, researchers are reporting today in a large study that could settle more than a half-century of scientific debate about the relationship between I.Q. and birth order. The average difference in I.Q. was three points higher in the eldest child than in the closest sibling. This paper is one of the most important publications to come out in this field in many years, researchers said.

A study of American child care has found that keeping a preschooler in a day care center for a year or more increased the likelihood that the child would become disruptive in class -- and that the effect persisted through the sixth grade. Every year spent in such centers for at least 10 hours per week was associated with a 1 percent higher score on a standardized assessment of problem behaviors completed by teachers. With more than two million American preschoolers attending day care, the increased disruptiveness very likely contributes to the load on teachers who must manage large classrooms, the researchers said.

Biased samples

A recent study shows that children enjoy playing educational video games just as much as they enjoy playing non-educational video games. The researchers conducted their study at a local university with children of university professors. They required the children to play a variety of educational and non-educational video games. They then asked children to rate how much they enjoyed playing the two types of games. Interestingly, children's enjoyment of educational video games was not significantly different from their enjoyment of non-educational video games.

A recent study shows that Americans are more interested in politics than was previously thought. Researchers approached people at events such as town meetings and city council meetings and surveyed them about their interest in politics and their voting behavior. Seventy percent of those surveyed reported that they were planning to vote in upcoming local and national elections. More than half said that they regularly read articles in the newspaper about political issues.

Appendix D

Actively Open-Minded Thinking Questionnaire

This questionnaire lists a series of statements about various topics. Read each statement and decide whether you agree or disagree with each statement as follows:

1 - Disagree Strongly, 2 - Disagree Moderately, 3 - Disagree Slightly, 4 - Agree Slightly, 5 - Agree Moderately, 6 - Agree Strongly

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. Be sure the number on the answer sheet corresponds to the number of the statement to which you are responding. There is no time limit, but work as quickly as possible.

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. (Reflected)

2. What beliefs you hold have more to do with your own personal character than the experiences that may have given rise to them. (Reflected)

3. I tend to classify people as either for me or against me. (Reflected)

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. (Reflected)

6. Changing your mind is a sign of weakness. (Reflected)

7. I believe we should look to our religious authorities for decisions on moral issues. (Reflected)

8. I think there are many wrong ways, but only one right way, to almost anything. (Reflected)

9. It makes me happy and proud when someone famous holds the same beliefs that I do. (Reflected)

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. (Reflected)

12. Abandoning a previous belief is a sign of strong character.

13. No one can talk me out of something I know is right. (Reflected)

14. Basically, I know everything I need to know about the important things in life. (Reflected)

15. It is important to persevere in your beliefs even when evidence is brought to bear against them. (Reflected)

16. Considering too many different opinions often leads to bad decisions. (Reflected)

17. There are basically two kinds of people in this world, good and bad. (Reflected)

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. (Reflected)

20. Most people just don't know what's good for them. (Reflected)

21. It is a noble thing when someone holds the same beliefs as their parents. (Reflected)

22. Coming to decisions quickly is a sign of wisdom. (Reflected)

23. I believe that loyalty to one's ideals and principles is more important than "open-mindedness." (Reflected)

24. Of all the different philosophies which exist in the world there is probably only one which is correct. (Reflected)

25. My beliefs would not have been very different if I had been raised by a different set of parents. (Reflected)

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. (Reflected)

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. (Reflected)

32. I believe that the "new morality" of permissiveness is no morality at all. (Reflected)

33. One should disregard evidence that conflicts with your established beliefs. (Reflected)

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. (Reflected)

36. Often, when people criticize me, they don't have their facts straight. (Reflected)

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. (Reflected)

39. I believe letting students hear controversial speakers can only confuse and mislead them. (Reflected)

40. Intuition is the best guide in making decisions. (Reflected)

41. People should always take into consideration evidence that goes against their beliefs.

Appendix E

Need for Cognition Questionnaire

For each of the statements below, please indicate to what extent the statement is characteristic of you. If the statement is extremely uncharacteristic of you (not at all like you) please mark a “1”; if the statement is extremely characteristic of you (very much like you) please mark a “5”. 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. There is no time limit, but work as quickly as possible.

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun. (Reflected)
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. (Reflected)
5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something. (Reflected)
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to. (Reflected)
8. I prefer to think about small, daily projects to long-term ones. (Reflected)
9. I like tasks that require little thought once I’ve learned them. (Reflected)
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn’t excite me very much. (Reflected)
13. I prefer my life to be filled with puzzles that I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult and important to one that is somewhat important but does not require much thought.

16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort. (Reflected)

17. It's enough for me that something gets the job done; I don't care how or why it works. (Reflected)

18. I usually end up deliberating about issues even when they do not affect me personally.

Appendix F

Epistemic Belief Inventory

This questionnaire lists a series of statements about various topics. Read each statement and decide whether you agree or disagree with each statement. 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. There is no time limit, but work as quickly as possible.

Responses:
1-strongly disagree
2-disagree
3-neutral
4-agree
5-strongly agree

1. Most things worth knowing are easy to understand.

2. What is true is a matter of opinion.

3. Students who learn things quickly are the most successful.

4. People should always obey the law.

5. People's intellectual potential is fixed at birth.

6. Absolute moral truth does not exist.

7. Parents should teach their children all there is to know about life.

8. Really smart students don't have to work as hard to do well in school.

9. If a person tries too hard to understand a problem, they will most likely end up being confused.

10. Too many theories just complicate things.

11. The best ideas are often the most simple.

12. Instructors should focus on facts instead of theories.

13. Some people are born with special gifts and talents.

14. How well you do in school depends on how smart you are.

15. If you don't learn something quickly, you won't ever learn it.

16. Some people just have a knack for learning and others don't.

17. Things are simpler than most professors would have you believe.

18. If two people are arguing about something, at least one of them must be wrong.

19. Children should be allowed to question their parents' authority.

20. If you haven't understood a chapter the first time through, going back over it won't help.

21. Science is easy to understand because it contains so many facts.

22. The more you know about a topic, the more there is to know.

23. What is true today will be true tomorrow.

24. Smart people are born that way.

25. When someone in authority tells me what to do, I usually do it.

26. People shouldn't question authority.

27. Working on a problem with no quick solution is a waste of time.

28. Sometimes there are no right answers to life's big problems.

Appendix G

Study 2 Background/Follow-Up Questions

Please answer the following questions:
1. What is your gender?
Male
Female

2. What is your age?

3. Is English your native language?
Yes
No

4. What year are you in at the U of M?
Freshman
Sophomore
Junior
Senior

5. What is your major?

6. What statistics and research methods courses have you taken and what were your grades in those courses? Please list each course and its grade on a separate line.

7. Did you find your statistics and/or research methods courses to be interesting and useful? Please answer separately for each course.

8. What has been your favorite course in college or high school?

9. What has been your least favorite course in college or high school?

10. Please report your SAT scores (n/a if you didn't take it).
Composite SAT score
Verbal SAT score
Math SAT score

11. Please report your ACT scores (n/a if you didn't take it).
Composite ACT score
English ACT score
Math ACT score

Note: The following items are all reversed for scoring purposes.

<u>Interest Items:</u>

1. How frequently do you read/watch/listen to news media outlets such as newspapers, magazines, TV news or radio news?
Every day
Multiple days a week
Once a week
Once or twice a month
Less than once a month
Never

2. I am interested in reading about science-related news.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

3. I am a critical reader of scientific studies.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

<u>Prior Belief Items:</u>

4. "When mothers' exert control over their children's eating this can cause children to overeat."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

5. "Scientists' creative accomplishments increase from ages 20-35 and then begin to

decline."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

6. "1 in 4 adolescents experience abuse every year."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

7. "Physical exercise, such as running, causes improved physical functioning in older adults."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

8. "Older siblings are more intelligent than younger siblings."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

9. "Children like to play educational video games just as much as non-educational video

games."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

10. "Children who attend daycare are more disruptive than children cared for at home."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

11. "More people are interested in politics than is usually thought."
Prior to completing this study, the above statement would've fit with my beliefs and experiences.
Agree strongly
Agree
Agree somewhat
No opinion
Disagree somewhat
Disagree
Disagree strongly

Familiarity with Statistics/Methodology Items:

12. Are you familiar with the idea that just because two variables are correlated doesn't mean that one causes the other?
Very familiar
Familiar
Somewhat familiar
Not very familiar
Not at all familiar

13. Are you familiar with the idea that one must clearly define a variable in order to measure it? This process is often know as operationalization.
Very familiar

Familiar
Somewhat familiar
Not very familiar
Not at all familiar

14. Are you familiar with the idea that you must examine the size of a quantitative research effect (a.k.a. effect size) in order to determine how large or important the result is?
Very familiar
Familiar
Somewhat familiar
Not very familiar
Not at all familiar

15. Are you familiar with the idea that, for the purposes of research, one must select a sample of participants that represents the population of interest?
Very familiar
Familiar
Somewhat familiar
Not very familiar
Not at all familiar

References

Andreoli, V. & Worchel, S. (1978). Effects of Media, Communicator, and Message Position on Attitude Change. *The Public Opinion Quarterly, 42* (1), 59-70.

Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.

Baron, J. (1987). *Being disposed to thinking: A typology of attitudes and dispositions related to acquiring and using thinking skills*. Boston, MA: University of Massachusetts, Critical and Creative Thinking Program.

Baron, J. (2000). *Thinking and deciding (3rd ed.)*: Cambridge University Press.

Baron, J., Badgio, P. C. & Gaskins, T. W. (1986). Cognitive style and its improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 173-220). Hillsdale, NJ: Erlbaum.

Brem, S. K. & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604.

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

Chandler, M., Boyes, M., & Ball, L. (1990).Relativism and stations of epistemic doubt. *Journal of Experimental Child Psychology, 50*, 370–395.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory. Odessa, FL: Psychological Assessment Resources.

Dawes, R. (1994). *House of cards: Psychotherapy built on myth.* New York: Free Press.

Edwards, K. & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments.*Journal of Personality and Social Psychology, 71*(1), 5-24.

Epstein, S., & Meier, P. (1989). Constructive thinking: a broad coping variable with specific components. *Journal of Personality and Social Psychology, 57*, 332–350.

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

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains:

Dispositions, skills, structure training and metacognitive monitoring. *American Psychologist, 55*(4), 449-455.

Halpern, D. F. (2001). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education, 50*(4), 270-286.

Halpern, D. F. (2007).The nature and nurture of critical thinking. In R. J. Sternberg, H. L. Roediger III, & D. F. Halpern (Eds.), *Critical thinking in psychology.* (pp. 1-14). New York: Cambridge University Press.

Harp, S. F., & Mayer, R. E. (1998). How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning. *Journal of Educational Psychology, 90*, 100-103.

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

Jehng, J. J., Johnson, S. D. & Anderson, R. C.(1993). Schooling and students' epistemological beliefs about learning. *Contemporary Educational Psychology, 18*, 23-35.

King, P. M., & Kitchener, K. S. (2002). The Reflective Judgment Model: Twenty years of research on epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing.* (pp. 37-61). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

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*(2), 89-116.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71*(5), 1347-1366.

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, 15*(3), 400-416.

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

Kuhn, D. (1991). *The skills of argument*: Cambridge University Press.

Kuhn, D. (1993). Thinking as an epistemological enterprise. *Applied Psychology: An International Review, 42*(3), 226-228.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science,6*, 133-139.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2),16-25.

Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1-8.

Kuhn, D., & Dean, D., Jr. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into practice, 43*(4).

Kuhn, D., Pennington, N., & Leadbeater, B. (1983). Adult thinking in developmental perspective: The sample case of juror reasoning. In P. Baltes & O. Brim (Eds.), *Life-span development and behavior, Vol. 5*(pp. 158–195). New York: Academic Press.

Kuhn, D., Weinstock, M. & Flaton, R. (1994). Historical reasoning as theory-evidence coordination. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences*(pp. 377-401). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5*, 289–296.

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). Longitudinal study of the effects of Undergraduate training on reasoning. *Developmental Psychology, 26*(6), 952-960.

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

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*, 2098-2109.

Meehl, Paul E. (2002). Cliometicmetatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. *Psychological Reports, Vol 91*(2), 339-404.

Messer, W. S., & Grigges, R. A. (1989). Student belief and involvement in the paranormal and performance in introductory psychology. *Teaching of Psychology, 16,* 187-191.

Miller, K. (1997, July). Star struck: A journey to the new frontiers of the zodiac. *Life,* pp. 39-44, 46, 50.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307–317.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly, 39*(1), 1-21.

Perry, W. G., Jr. (1970).*Forms of intellectual and ethical development in the college years*. Oxford, England: Holt, Rinehart & Winston.

Petty, R. E., Priester, J. R., & Brinol, P. (2002). Mass media attitude change: Implications of the Elaboration Likelihood Model of persuasion. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 155-198). Hillsdale, NJ: Erlbaum.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research, 42*(3), 237-249.

Priester, J. M., & Petty, R. E. (1995). Source attributions and persuasion: Perceived honesty as a determinant of message scrutiny. *Personality and Social Psychology Bulletin, 21,* 637-654.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991*).Measures of personality and social psychological attitudes* (Vol. 1). San Diego, CA: Academic Press, pp. 560–5 64.

Sá, W., Kelley, C., 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.

Schommer, M., & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology, 87*(3), 424-432.

Schwartz, D. L. & Bransford, J. D. (1998).A time for telling. *Cognition and Instruction, 16*(4), 475-522.

Solomon, G. & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychology, 24*(2), 113-142.

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.

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*(7), 851-860.

Troldahl, V., & Powell, F. (1965).A short-form dogmatism scale for use in field studies. *Social Forces*, *44*, 211–215.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*(1), 45-69.

Weinstock, M. P., Neuman, Y., & Glassner, A. (2006). Identification of Informal Reasoning Fallacies as a Function of Epistemological Level, Grade Level, and Cognitive Ability. *Journal of Educational Psychology, 98*(2), 327-341.

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*, 470-477.

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

Zimmerman, C. (2005). *The Development of Scientific Reasoning Skills: What Psychologists Contribute to an Understanding of Elementary Science Learning.* Washington, D. C.: National Research Council Committee on Science Learning Kindergarten through Eighth Grade.