Improving Causal Reasoning in a College Science Course

Michael Harrington

University of Michigan

Mentors:

Prof. Colleen Seifert

Prof. Priti Shah

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of Bachelor of Science

With Honors in Psychology from the

University of Michigan 2019

Abstract

People often make analytical errors when reasoning about scientific findings, affecting their ability to make well-informed decisions in everyday life. A specific error often noted is making a causal claim from correlational evidence. This study investigated the thought processes involved when students reason about causal claims found in news media reports. What do students notice, or fail to notice, about causal claims in scientific studies? Can analytical thinking be improved by instruction clarifying the nature of causal versus correlational evidence? We examined changes in participants' performance on a series of three questionnaires one week apart. Each questionnaire asked students to critique a short media story about a scientific study by rating its quality and support for a causal claim, and then stating their reasons for their ratings. They were also given a second claim from an unrelated study and were asked to create diagrams depicting its possible causal relationships. At the beginning of session 2, students received an instructional intervention that identified the correlational/causation error through an extended example with questions to elicit their thinking about alternative causes. They then completed the same two measures with different stories both immediately and 1 week later. The results showed improvement on both measures of students' reasoning about causal relationships after the intervention, and that improvement was maintained one week later. However, ratings of study quality did not follow this pattern. The results suggest interventions aimed at the causation vs. correlation error in reasoning are successful even when set in a larger course context of learning about evaluating scientific claims.

Improving Causal Reasoning in a College Science Course

Accessibility of scientific information has never been better; as digital media has expanded, people have become regular consumers of scientific claims from research journals, television, magazines, blogs, or even word of mouth (Bromme & Goldman, 2014; Baram-Tsabari & Osbourne, 2015). However, applying scientific findings in everyday life also requires an evaluation of the scientific claims presented. To make well-informed decisions based on science, it is important for individuals to make distinctions between “good” science, “bad” science, appropriate claims, and pseudoscience (Kolsto, Bungum, Arnesen, Isnes, Kristensen, Mathiassen, & Ulvik, 2006; Trefil, 2008). Media stories containing scientific claims often misrepresent, oversimplify, or overdramatize the implications of their findings (Bromme & Goldman, 2014). Individuals must be able to understand scientific claims and limitations through “scientific literacy” skills (Durant, 1993, p.129), which imply “… an appreciation of the nature, aims, and general limitations of science, coupled with some understanding of more important scientific ideas” (Jenkins, 1994, p.5345). This ability is integral when it comes to using scientific findings to guide decisions regarding health, behavior, and public policy (Kolsto et al., 2006; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012).

Because many media articles present scientific findings from an oversimplified perspective, people may not carefully consider casual explanations, and may inaccurately interpret scientific results (Bromme & Goldman, 2014; Stadtler, Scharrer, Brummernhenrich, & Bromme, 2013). In addition, the reader may assume a source discussing science provides authoritative information no matter where it is encountered (Yeo & Tan, 2010), or assume needed information is simply omitted in the abbreviated media report. Furthermore, the tendency to rely on one’s own personal beliefs and experiences when considering scientific data is exacerbated by certain features of media stories, such as use of personal anecdotes. Anecdotal evidence increases experiential thinking and decreases deep, analytic thinking (Rodriguez, Ng, & Shah, 2016; Rodriguez, Rhodes, Miller, & Shah, 2016). While scientific claims are now readily available to the public, comprehension of such claims is still often challenging (Bromme & Goldman, 2014).

Research has shown that students -- and even some trained scientific professionals -- make errors when reasoning about scientific evidence (Halpern, 1998). These errors don't apply only to students and professionals: More than 70% of American adults hold at least one pseudoscientific belief (Moore, 2005). Furthermore, one-third of Americans think that evolution is "absolutely false," and another 21% are unsure (Miller, Scott, & Okamoto, 2006). Less than half of Americans believe that the world is billions of years old (Bishop, Thomas, Wood, & Gwon, 2010), and only about 30% of American adults are able to understand stories in the science section of the New York Times (Anelli, 2011). While scientific reasoning skills are typically acquired through expository texts about natural science (Schwichow, Croker, Zimmerman, Höffler, & Härtig, 2016), similar scientific claims made in media articles require people to apply their reasoning skills wherever they arise.

Perhaps the most common error in causal reasoning involves *theory-evidence coordination*. A correlation (the *evidence*) between two variables does not imply causation (the *theory*), but people often interpret correlational findings as supporting causal claims (Burrage, 2008; Hatfield, Faunce, & Job, 2006; Rodriguez, Ng, & Shah, 2016; Rodriguez, Rhodes, Miller, & Shah, 2016). When two variables *A* and *B* are correlated, the association may reflect one or more causal relationships between them: *A causes B*; *B causes A; a third variable causes both A and B*; it is bidirectional (*A causes B* and *B causes A*); or, the correlation may be spurious (without any relationship) (Shah et al., 2017). Unfortunately, people rarely generate alternate mechanisms for such causal alternatives even when explicitly asked to do so (Shah et al., 2017).

For example, consider this headline about a psychology study: “Smiling promotes longevity” (WorldHealth, 2010). Without more information, what might a reader reasonably conclude? The headline appears to assert that smiling *causes* longevity. How might smiling more cause a longer life span? While a causal mechanism between smiling more and living longer may not be evident, other possible causal relationships are. Living longer would logically afford more opportunity to smile, and living a happy life might account for more of both smiling and years of life. Generating these alternative causal relationships may help the reader recognize that the specific causal relationship is not sufficiently documented, and should be rejected. In fact, the journal article describing the study is entitled, “Smile intensity in photographs predicts longevity” (Abel & Kruger, 2010), employed a keyword (predicts) reserved for correlation.

Alternatively, suppose the headline read, “Smiling promotes attractiveness.” In this case, smiling as the cause may be more plausible, and a third variable causing both appears implausible. Further, an experimental study where people are assigned to either smile or not may identify the effect on attractiveness; indeed, this claim comes from an experimental study (Golle, Mast & Lobmaier, 2014). While far from certain, it may be reasonable to conclude that this study may indeed support a causal claim. Of course, reading the scientific article is the best way to evaluate the evidence provided; however, media reports rarely provide even a link to the journal. In such cases, people are left on their own to consider the plausibility of a causal inference. In everyday circumstances, people do apply causal reasoning to go “beyond the information given” to make their own conclusion (Waldmann, Hagmayer, & Blaisdell, 2006).

Gaining skills to critically evaluate scientific evidence is a major goal of K-12 education (Lehrer & Schauble, 2006; Next Generation Science Standards (NGSS)). However, teaching scientific content alone does not appear to improve students' reasoning about science (Crowell & Schunn, 2016). While Chinese students learn a more science content, research has shown that they perform equally to U.S. students on measurements of scientific reasoning (Crowell & Schunn, 2016). Additionally, those taking at least eight college science classes did not outperform high school students on scientific reasoning tasks (Norris & Phillips, 1994; Norris, Phillips, & Korpan, 2003), though subsequent studies document a correlation between college science training and science evaluation skills (Amsel et al., 2008; Burrage, 2008; Huber & Kuncel, 2015; Kosonen & Winne, 1995; Norcross, Gerrity, & Hogan, 1993).

What skills do students need in order to understand scientific claims and to avoid errors in reason about causes? They must be able to identify threats to scientific validity (Anelli, 2011; Miller 1996, Picardi & Masick, 2013; Reis & Judd, 2000; Sagan, 1996a) and judge whether claims provide adequate support for a conclusion, referred to as *theory-evidence coordination* (Kuhn, Amsel, O'Loughlin, Schauble, Leadbeater, & Yotive, 1988). In other words, providing individuals with more knowledge about how to critically evaluate scientific evidence does not guarantee that they'll be able to apply those skills in everyday contexts (Shah, Michal, Ibrahim, Rhodes, & Rodriguez, 2017). College students performed no better than 7th or 10th grade students when evaluating the validity of an experimental study; however, they performed better when explicitly asked *think critically* about the study (Kosonen & Winne, 1995; Rodriguez, Ng, & Shah, 2016; Rodriguez, Rhodes, Miller, & Shah, 2016). It appears that people have the ability to engage in analytic thinking, but tend not to do so unless specifically prompted.

One contributing factor may be that people generally engage in heuristic rather than analytic thinking, because it is easier. Heuristic thinking is characterized as fast, frugal, automatic, emotional, and unconscious (Kahneman, 2011). When analytic thinking is applied, individuals are more likely to recognize threats to scientific validity (Shah et al., 2017). However, analytic thinking requires a substantial amount of effort, and due to limited cognitive resources, occurs less frequently than heuristic thinking (Shah et al., 2017). Due to its ease, heuristic thinking can quickly lead people through tasks at hand, but often results in reasoning errors (Shah et al., 2017). Studies show that heuristic thinking is predominant in K-12 students, college students, and the lay public when reading media articles containing scientific content (Norris & Phillips, 1994).

The ability to understand scientific claims and evaluate evidence appropriately may further be limited by personal beliefs, prior experiences, and emotional responses (Shah et al., 2017). When evidence "feels right" (often when congruent with one's prior beliefs) or makes one feel good (because it is desired), people are more likely to judge the evidence as "high quality," without considering other factors (Shah et al., 2017). It is only when presented with information that goes against their existing beliefs that individuals stop engaging in heuristic thinking and begin to engage in more analytic, critical thinking (D. Evans, 2003; J. Evans, 2003; Evans & Curtis-Holmes, 2005; Klaczynski, 2000; Kunda, 1990; Sà, West, & Stanovich, 1999; Sinatra, Kienhues, & Hofer, 2014; Nickerson, 1998).

Understanding scientific findings may also be influenced by the presence of visualizations; for example, a scatterplot can alter the conclusions drawn from scientific evidence. In one study, Ibrahim, Seifert, Adar, & Shah asked people to read about the safety of genetically modified organisms (GMOs) (2016). Following information describing decades of research supporting the conclusion that GMOs are safe, *one* fictional study reported a positive correlation. The results showed that people were more likely to adopt the new evidence and make further causal inferences if a visualization was included (Ibrahim, Seifert, Adar, & Shah, 2016). Other visualizations have also been shown to encourage adoption of new evidence, including bar graphs (Tal & Wansink, 2016), scientific formulas (Tal & Wansick, 2016), and neuroscience information (Rhodes, Rodriquez, & Shah, 2014). These studies demonstrate how easily people can be swayed by presentation features without regard to merit (Shah et al., 2017).

However, visual explanations, when generated by the *learner*, can improve their ability to understand scientific data (Bobek & Tversky, 2016; Gobert & Clement, 1999; Ainsworth & Loizou, 2003). Creating diagrams after reading scientific material has been linked with better understanding of causal and dynamic relationships (Gobert & Clement, 1999). This suggests diagramming may motivate people to consider alternative causal relationships, a key to reasoning about causal claims in science. Further, participants viewing diagrams generated more self-explanations (Ainsworth & Loizou, 2003) to explain the rationale of example solutions to themselves (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Studies of example-based instruction (Shafto, Goodman, & Griffiths, 2014; Van Gog & Rummel, 2010) show improvements in learning through spontaneous, prompted, and trained self-explanations of examples (Renkl, Stark, Gruber, & Mandl, 1998).

In addition, examples that are *incorrect* have been shown to be beneficial to learners, perhaps because they help them avoid these errors later (Siegler & Chen, 2008). For example, Durkin and Rittle-Johnson (2012) found that studying common mathematical errors facilitated learning even for students with limited prior domain knowledge. Students' explanations during the intervention revealed that those in the *incorrect* condition more frequently discussed correct concepts. This suggests learning from examples of errors may be an effective intervention for supporting reasoning about causal relationships; because the theory-evidence coordination error is so prevalent, learning to recognize this error may show improvements in causal reasoning performance. Other studies have shown that college students rarely notice this error when it occurs in others' thinking (Burrage, 2008; Rodriguez, Ng, & Shah, 2016; Rodriguez, Rhodes, Miller, & Shah, 2016); if brought to their attention through instruction, learners may benefit from understanding why this error occurs so often.

The present study addresses this possibility by investigating whether causal reasoning in "everyday science" contexts can be improved through an extended example of the theory-evidence coordination error. By enhancing understanding of the error, and how to consider alternative causes, reasoning about future claims may be improved. The intervention for the study (see Appendix A) describes an extended example of this error in a real-life context: In 20 states, graduation requirements were raised to include Algebra II because a study showed Algebra II predicts college and work success (Carnevale, Strohl, & Smith, 2009). This fits the pattern of an *A causes B* claim arising from correlational evidence, and alternative causal relationships are readily apparent. As Anthony Carnevale said about the study, "The causal relationship is very, very weak. Most people don't use Algebra II in college, let alone in real life. The state governments need to be careful with this" (Whoriskey, 2011).

To enhance learning through the intervention, additional support was provided by explicitly describing the error made (by legislators in this instance) (Große & Renkl, 2007) and providing an explanation about why certain causal inferences (*A causes B*) were not correct given the evidence (Stark, Kopp, & Fischer, 2011). Further, following Berthold and Renkl (2009, 2010), prompts were inserted to encourage students to "work through" the reasoning about alternative causal models. Students were asked if specific alternative causes might account for the observed association between Algebra 2 and better jobs, including whether students already taking Algebra 2 were smarter, already headed to college, and in better high schools. The learning outcomes may be superior if students receive help in this format where an instructional explanation asks learners to understand the theory-evidence correspondence error and participate in the causal reasoning steps required to assess the causal claim.

Finally, the intervention included a series of causal diagrams illustrating the multiple possible relationships related to the observed association.

The study examines understanding of “everyday” media articles referring to scientific studies; typically, a brief description of correlational evidence is presented along with a headline claiming a causal relationship:

> *Although some people prefer to work in silence, many people opt to listen to music while working or studying. A recent survey study was conducted at a large midwestern university that found that students who had listened to music while studying received higher test scores than those who didn’t. The research team concluded that students who want to do well on exams should study while listening to music.*

Understanding this story and its implications for a causal claim requires some knowledge about theory-evidence coordination, as well as causal reasoning skills to identify potential alternative theories. Furthermore, the reader must also understand which claims are appropriate based upon assessing the type of evidence collected in the study (e.g., a test of association, an experimental design). Critically, reasoners must know that only experimental designs can support claims of causal relationships between variables.

The study was conducted in a college research methods course through a series of three in-class sessions. The pretest included a questionnaire with two measures of causal reasoning: a “study critique” task where a brief scientific report made a causal claim and students rated its appropriateness and explained their reasoning; and a “causal diagram” task where students generated causal diagrams for given a new *A causes B* claim. Session 2 began with the intervention task, including the extended example of theory-evidence correspondence error, additional questions and visualizations to guide students in considering alternative causes, followed by the same two measures. Finally, a posttest the following week again repeated the two measures. We expected that the example and explanation of an important “correlation is causation” error would improve students’ ability to recognize and understand how correlational evidence corresponds to theoretical causes, and assist them in avoiding the theory-evidence coordination problem in their own reasoning when analyzing scientific claims.

Method

Participants

Students enrolled in a large midwestern university course on Psychology research methods were invited to participate in this study. All students attending the course lecture for three consecutive weeks took part in the study. Of the 258 enrolled, 97 (37%) completed all three and were included in the analysis.

Materials

The two measures of causal reasoning -- Study Critique and Causal Diagram -- were included in each session as a repeated measure. Hardcopy questionnaires were prepared with *pretest, intervention,* or *posttest* tasks. Each task set included two problems selected from brief news reports of actual studies (see Appendix B). In order to avoid item-specific effects, three separate problem sets were created. Each set contained one Study Critique problem and one Headline Diagram problem paired at random. The three problem sets were then counterbalanced across students by assigning them at random to one of three alternate forms. The three sets were designated by colored cover sheets (green, white, or yellow) to facilitate matching students with their form in each session. A schematic illustrating the questionnaire content for each session is shown in Figure 1.

Study Critique Task. To assess students' ability to evaluate a causal claim, a short media story (between 100-150 words) describing a study (see Appendix A) was presented followed by questions. Two rating items assessed the perceived "quality" of the presented study and whether the findings "support the claim." These questions used a 5 point Likert scale, with "1" indicating low-quality and an unsupported claim, and "5" indicating high-quality and a claim supported by the study. Next, an open-ended question asked for "a critical evaluation of the study's method and claims, and what was good and bad about it" (see Figure 2). Examples of students' critical evaluation responses are shown in Figure 3.

Diagram Instructions. A one-page sheet (see Figure 4) gave instructions with three examples of causal diagrams, and then asked students to draw their own diagram to show, "Hyperactivity causes a faster metabolism." Ninety-eight percent of students completed this sheet with correct diagrams, suggesting they understood the instructions about how to make causal diagrams.

Casual Diagram Task. A second type of problem presented a short "headline" claim such as "Smiling increases longevity," and asked students to "diagram possible relationships between variables suggested by" the news headline. Two examples of the question with student answers are shown in Figure 5.

Intervention. The intervention was a text-based guide to understanding causal claims through an extended example of a correlation/causation error in reasoning. The example involved a news story about a link between taking advanced algebra courses in high school and later career earnings. Lawmakers assumed this was a causal link, and now require all students in the state to take advanced algebra in high school. Short exercises followed in a worksheet format to engage students in generating and considering alternative causes (see Appendix B). The intervention booklet also presented alternative causal models for the observed association, and ended with written advice on considering a causal claim when correlational evidence is presented. Several questions assessed self-rated understanding of the intervention content.

Procedure

The study took place during lecture sessions for the advanced psychology course. At the beginning of the lecture for each week of the study, students were asked to complete a questionnaire. For session 1, participants were randomly assigned to one of three forms and asked to complete the questionnaire. In the following week's session, students were given the matching booklet (green, white, or yellow) from the first week. In session 2, they completed the written intervention worksheet followed by the second questionnaire to assess changes in reasoning. In the third week, students again received the appropriate booklet (green, white, or yellow) completed the posttest as a third repeated measure to document any improvements in understanding causal relationships.

Measures

Ratings. The two rating items indicated a correct assessment of the study quality and support measures (e.g., that the study is not of high quality and the results do not support the claim) through lower scores. These two ratings were added together for each study critique problem as a *Ratings* score ranging from 2 to 10. An decrease in ratings over time reflect improved understanding of the problems with making causal claims from correlational data.

Reasons. Students' written responses in the open-ended study critique were qualitatively coded. Six codes captured as emergent themes were scored as the *Reasons* measure: 1) correlation is not the same as causation; 2) other variables may potentially play a role in the observed correlation; 3) additional information is needed to make a causal claim; 4) the study describes a survey rather than an experiment; 5) the study was not a randomized controlled trial to establish causation; and 6) additional studies are needed to be conclusive about a causal claim. The more codes identified in the critique (out of six), the higher the score on the *Reason* variable. Two raters independently coded ten percent of the responses, and the percentage of agreement between scores was above 92 percent. The first author then scored the remainder of the responses.

Themes. The *Themes* measure captured how well students could generate their own alternative causal relationships between the variables. The Causal Diagram task offered a text headline (e.g., "Breast-feeding increases intelligence") asserting an *A*->*B* causal claim, and students illustrated possible causal relationships through one or more drawings linking variables. Responses were coded using deductive qualitative codes capturing separate causal themes: 1) a standard *A* -> *B* diagram; 2) a *B* -> *A* reverse direction diagram; 3) two or more variables shown to cause one other; 4) one "third variable" shown to cause changes in two or more other variables; 5) linked variables through multiple steps; and 6) connecting variables not stated in the claim.

Two raters independently coded ten percent of the responses, and the percentage of agreement between coders was above 90 percent. The first author then scored the remainder of the responses. A higher number of codes identified in a student's response indicated more ability to generate alternative causal scenarios. An increase in the ability to generate these diagrams over sessions reflects an increased ability to think about alternative causal explanations for an observed association between variables.

Results

In this repeated-measures design, each student completed three different questionnaires, each including the same six test problems appearing in three different orders. Approximately equal numbers of students completed each (n = 30, 32, 34). A repeated measures analysis with *Form* (green, white, or yellow) as a between-groups factor found no main effects, with no significant differences for *Ratings* ($F(2, 94) = 0.996$, $p = 0.373$), *Reasons* ($F(2, 93) = 2.64$, $p = 0.08$), or *Themes* ($F(2, 93) = 1.84$, $p = 0.167$). No further comparisons were made by form groups.

Several questions within the intervention allow some determination of its success. For example, when the intervention asked whether the example (incorrectly arguing causation from a correlation) is "convincing," some students (35%) said, "yes," and their stated reasons were wrong for 53% and partially correct for 42%. This suggests seeing the initial example on the intervention resulted in the sense that the study should not be convincing; however, they were unable to articulate why. Later, they indicate whether four alternative causes may be responsible for higher earnings instead of taking Algebra; of these, two alternative variables ("smarter" and "richer") were endorsed as plausible by about half of the students, and about 75% endorsed two others ("better schools" and "headed to college anyway"). Then, when three alternative causal diagrams are presented to illustrate other possible causes, over 75% answered correctly. Based on these measures, it appears most students assessed their own reasoning as understanding the intervention, and most were able to reason about alternatives to the stated causal claim.

A repeated measures ANOVA showed significant effects of *Session* (pre-test, intervention, and post-test) on causal reasoning as measured on all three dependent variables (*Ratings, Reason*, and *Themes*). One measure of causal understanding from the Study Critique task is *Ratings* (ranging from 2 to 10), combining students' assessment of the study's quality and its support for a causal claim. The ratings suggest a moderate level of understanding, with 34% giving it a low-quality rating (2 or 1 on the scale) at pre-test, 40% after the intervention, and 26% on the post-test.

The rating of support for the causal claim in the study was similarly midrange, with 48% correctly saying it was of low quality at the pre-test, 40% after the intervention, and 37% at the post-test. This suggests the majority of students did not recognize the study's evidence as a poor fit for a causal claim. A planned linear contrast shows a significant change in *Ratings* scores over the three sessions, $F(1, 94) = 7.27$, $p < 0.008$, $\eta^2_p = .072$, with better performance at pre-test (M = 6.47, SD = 1.671) and intervention (M =6.397, SD = 1.732) than at post-test (M =7.01, SD = 1.636). This linear trend in ratings is shown in Figure 6.

Rather than downgrading their ratings of the study's quality and support for a causal claim, students *increased* their approval of the study over the three sessions (with high ratings = "poor quality"/ "good support for claim"). On the pretest ratings, students averaged 6.47, above the midpoint of the twelve-point combined ratings scale, indicating they did consider the study to provide good support for a causal claim, and their support increased by the post-test. The trend shows students were *more* accepting of the study's quality and support for a causal claim at the end of the study. One reason for this finding may be the rating scale: If students felt the correlational study cannot support a causal claim, they should rate study quality as low (closer to "1") and support for claim as, "not at all." It is possible that students felt confusion about how to assess "quality" in studies where other factors may play a role; for example, on the pre-test, students mentioned the sample size of the study, endorsements by doctors, specialized populations, and bias in respondents, all factors that may cause concern about a study, but are minor in comparison to the causal claim made. Consequently, ratings of quality and support may increase across session because they are not adequate measures of knowledge for capturing reasoning about causal claims.

A second measure of causal understanding from the Study Critique task was *Reasons*, scored as a count of coded themes related to causality (range is 0 to 6). The planned linear contrast indicating improvement in *Reasons* scores over the three sessions was significant, $F(1, 94) = 9.318$. $p < 0.003$, $\eta^2_p = .090$. This qualitative measure reflects students' concerns about the study's correlational evidence as support for a causal claim. This measure shows a substantial improvement from pretest (M = 1.25; SD = .854) to just after the intervention (M = 1.56, SD = .841), and this improvement is maintained a week later (M = 1.58, SD = .875). These findings show an increased ability to discuss reasons related to theory-evidence correspondence in students' open-ended responses critiquing the presented study. In their own words, students increased the number of reasons related to the appropriate use of correlational evidence immediately after seeing the intervention material, and maintained their improvement on the post-test one week later. (See Figure 7.)

The causal diagrams task allowed students to create their own representations of possible causal relationships for a stated, *A causes B* headline. A planned linear contrast showed a significant increase in the number of causal relationships included in the diagram ($F(1,93)$ =

30.935, $p < 0.001$, $\eta^2_p = 0.25$) from pre-test (M = 1.68, SD = .985) to intervention (M = 2.44, SD = 1.429), and maintaining the gain at post-test (M = 2.40, SD = 1.469), as shown in Figure 8. The causal diagrams created by students show an increase in the number of causal themes evident in their diagrams (reverse causes, third variables, multiple diagrams), reflecting their increased ability to consider alternative causes for an observed correlation. As with the causal reasons for the critiques, students maintained their gains in the post-test measure one week later.

Discussion

The present study provides evidence about how well college students understand the requirements for making a causal claim from a correlational finding. Following the intervention with an extended example of this error in session 2, students increased their use of causal reasoning in critiquing studies with correlational data, and increased the types of causal relationships represented in their diagrams of possible relationships between two variables. These findings suggest improved causal understanding was evident when examining the qualities of students' alternative causal explanations for associations. However, while rating scales of study quality and support for the causal claim show change over the sessions, this measure does not correspond to improvement in causal reasoning, suggesting confusion about what elements are critical to scientific quality. These results show measures of causal reasoning can diverge based on how specifically causal concepts are referenced, with open-ended responses on two tasks indicating improvement over the sessions following the intervention.

Despite these reliable findings, an important result from the study is that the impact of the intervention (though statistically significant) is modest, adding on average less than one of six reasons (6%) for concerns about the study and less than one of six themes (12.5%) included in causal diagrams. While students offered a wide range of alternative causal models in their diagrams (as detailed in the six themes), their lower production of causal alternatives in the study suggests there is room for improvement in students' performance, especially at the level of college courses.

Across the study, students simply did not write many causal reasons or alternative themes in their responses, and appeared to write a minimal response. A low level of responses written may arise from the large-session administration of the measures in an (ungraded) class activity; consequently, students may have written progressively less over the sessions. In support of this possibility, the number of diagrams created also decreased over sessions despite reliable increases in the quality of their causal alternatives identified. While students did increase their correct use of causal understanding, they did not demonstrate a high degree of causal inference in their writing about each example study.

The ability to consider multiple causal models (as in the diagram task) demonstrates objective and creative thinking because students must posit additional variables (third variables not stated in the problem) and relationships (*B causes A*, multiple causes and multiple effects) in ways that frame the association as a pattern explainable by multiple causal relationships between the variables. The findings show students were able to add their own potential causal relationships to their responses after the intervention, demonstrating the effectiveness of the intervention in increasing their understanding of causal relationships.

The study employed measures where causal reasoning is applied to specific descriptions of science studies and claims. Applications of new knowledge outside the classroom may be less successful than direct measures of learning; for students, connecting classroom learning about correlations to the reasoning required to assess causal claims may be more challenging. Some prior studies have found students can make use of statistical knowledge gained in a class when asked about a principle through an unexpected phone call in another context (Fong, Krantz, & Nisbett, 1986). More evidence is needed to show the benefits of this intervention extend to causal reasoning in the target contexts such as reading news media where science summaries are presented. As a first step, this study demonstrated higher levels of performance on understanding causal claims within the classroom, providing a foundation for examining their spontaneous use of the learning in external settings.

As a classroom study, a major limitation is the attrition of students over three consecutive weeks of the study. As attendance in lecture was optional, approximately a third of the enrolled students completed all three sessions, and no information is available about differences in this subsample compared to the larger class. As the study continued through the three weeks, participants may have lost interest in the content of the study, resulting in less effort and lower completion over time. This design was employed in order to demonstrate the effects of the intervention as long-lasting, maintained over a week until the posttest session. Future studies may complete a pre-test, intervention, and post-test in just one session to examine the impact of interventions on immediate performance.

Students in the study are likely to represent able learners with a great deal of prior knowledge about experiments. As psychology majors, their studies may have prepared them to benefit from the specific intervention employed. While learning from error examples has been successful in other studies (Durkin & Rittle-Johnson, 2012; Siegler & Chen, 2008). Future studies should include a more diverse population of students at varied levels of science learning experience to determine whether more novice learners can benefit from science instruction with this type of intervention, and include comparisons to objective measures of academic performance such as test scores. As college students, the intervention materials may have been easier to understand due to more experience in the classroom context. For those not in college courses, the intervention helpful in this study may be less convincing or more difficult to understand.

The results from this study are promising in that students increased their consideration of alternative causal models when presented with correlated variables. The error example in the intervention included state legislatures enacting laws based on a theory-evidence correspondence error; potentially, the example showing adults (who should know better) making the error may be especially helpful in students' learning. It is also possible this example was more meaningful for the students in the study as recent high school graduates in a state where the Algebra 2 requirement was in place for them; in addition, other factors affecting their own interest in higher levels of mathematics may be more prominent based on their own experiences. The success of this intervention suggests a strategy for instruction based on identifying causal reasoning errors that "hit close to home" as helpful in engaging causal reasoning.

The results confirm prior findings showing the use of visualizations can improve science learning. As noted by Gobert & Clement, creating diagrams following reading scientific material has been linked with better understanding of the causal/dynamic aspects of the study (1999). In the causal diagram task in our study, students were asked to create their own drawings of possible causal relationships underlying the *A causes B* claim. Participants saw a model for how to diagram results in the intervention, and it had a positive effect on their later performance. Models of reasoning provided by worked examples may be especially important in learning to reason about new problems (Renkl, Stark, Gruber, & Mandl, 1998). Self-generated explanations encouraged during the intervention through visual explanations has been shown to improve learners' ability to understand scientific data (Bobek & Tversky, 2016; Gobert &

Clement, 1999; Ainsworth & Loizou, 2003; Ibrahim et al., 2016; Tal & Wansink, 2016; Rhodes, Rodriquez, & Shah, 2014).

Studies of example-based instruction (Shafto, Goodman, & Griffiths, 2014; Van Gog & Rummel, 2010) show improvements in learning through spontaneous, prompted, and trained self-explanations of examples (Ainsworth & Loizou, 2003; Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). In particular, the present study supports the role of examples illustrating error in learning to apply scientific reasoning. While learning from error has been identified in other complex settings (e.g., Seifert & Hutchins, 1990), the notion of instructing students about common errors in reasoning about science may hold promise. Because specific errors are well documented, teaching students about the error pattern – that making a causal claim from correlational data is in error, and explaining why – may be a helpful approach for other errors. Because students may struggle to recognize the error when it occurs in the work of others or in their own work, it is important to provide experience with error recognition in order to build skills for error detection. Further studies can extend this approach to science education by providing other instructional interventions based on common errors.

These results are also consistent with theories of cognitive processes in science education. For example, when asked to explain their reasons for why the study is convincing, students often invoked heuristic thinking in responses such as, "…because that's what I've heard before" or "…because that makes sense to me." As heuristic thinking is a quick alternative to careful, deliberate reasoning (Kahneman, 2011), students may easily gloss over the need to consider alternatives not stated in the problem or alternative relationships among the variables, leading to errors in reasoning (Shah et al., 2017). Furthermore, analytic thinking requires a substantial amount of effort, and due to limited cognitive resources, occurs less frequently than heuristic thinking (Shah et al., 2017). When considering scientific evidence, and when seeing reports in congruence with previous beliefs, students often rate a study as being of high quality, as seen in our study.

The study has implications for teaching about correlational studies in the psychology classroom, along with other science settings. Puzzling findings (such as the Algebra and salary association) may be helpful in encouraging people to think through other ways variables may be related. Further evidence relating these qualities of examples to the impact of interventions are needed. Often, the study described in a media report is more circumspect in its claims, but the media report highlights one possibility – the causal claim – over alternatives. It may be possible to train people to watch for claims of this form, and then reason about alternatives without engaging in a study critique. After all, the question of whether a correlational study is of "good quality" does not address its causal claim. A more apt question is whether people see the causal claim as "supported by" the study; in addition to promoting causal understanding, intervention efforts must address people's understanding of science studies.

This larger problem of assessing the quality of scientific studies is a more challenging agenda in science learning. Science studies in varied areas may appear quite different from one another, especially in the short summaries provided in the media. It appears to be unusual to find the needed information within a media article in order to recognize even whether the data were correlated or whether the study was a true experiment (with randomized assignment to groups). Only by engaging with a detailed description can a reader determine what support is offered by a given study for a causal claim. In everyday science contexts, the goal for readers informed by science is not necessarily to track down the needed information; instead, causal reasoning is invoked in order to assess and appreciate alternative accounts of possible causal influences.

Reasoning "in a vacuum of information" is often required when partial information is available; while postponing an assessment until information is complete is not incorrect, in practice, reasoning about the potential meaning of an association is a key skill gained from the study of science.

Science is used to advance recommendations in a wide variety of areas, and causes are sometimes impossible to assess; for example, associations such as smoking and health risks are simply unethical to test in experimental studies. Similarly, studies of child development differences are not experimental due to differences in families, gender, and other factors. While many understand the basics of experiments in science, the point of an experiment – to identify a cause for an effect – may not be prominent in science education. Addressing public science literacy may require changing the conversation to note that, "correlations are all around," but only a small number of news reports reflect true causal claims identified through science. That is, understanding the "base rate" for true causal claims may be important in raising skepticism about correlations, noting that the value of scientific enterprise may be inherent in the difficulty of establishing causal relationships. The underlying problem for research on causal reasoning is helping people to recognize, question, and apply their knowledge of causal relationships whenever relevant in everyday life. The present study provides some evidence for a small step in this direction.

Acknowledgments

I'd like to personally thank Professor's Shah and Seifert for their assistance and continued support throughout this study. I'd also like to thank my twin sister, Meghan, who also helped me through this process by assisting me in the coding of results, as well as by motivating me to perform at my best. I'd also like to thank my parents and family for their continued support throughout this process.

References

Abel, E. L., & Kruger, M. L. (2010). Smile intensity in photographs predicts longevity. *Psychological Science*, *21*(4), 542-544. doi: 10.1177/0956797610363775

Ainsworth, S., & Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*(4), 669-681. doi: 10.1207/s15516709cog2706_6

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., & Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. *Cognitive Development*, *23*(4), 452-471. doi: 10.1016/j.cogdev.2008.09.002

Anelli, C. M. (2011). Scientific literacy: What is it, are we teaching it, and does it matter? *American Entomologist*, *57*(4), 235-244. doi: 10.1093/ae/57.4.235

Baram-Tsabari, A., & Osborne, J. (2015). Bridging science education and science communication research. *Journal of Research in Science Teaching*, *52*(2), 135-144. doi: 10.1002/tea.21202

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*, 70–87. doi: 10.1037/a0013247

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22*, 25–40. doi: 10.1007/s10648-010-9124-9

Bishop, B., Thomas, R. K., Wood, J. A., & Gwon, M. (2010). Americans' scientific knowledge and beliefs about human evolution in the year of Darwin. *Reports of the National Center for*

Science Education, 30(3), 16-18. Retrieved from http://ncse.com/rncse/30/ 3/americans-scientific-knowledge-beliefs-human-evolution-year.

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications*, *1*(1). doi: 10.1186/s41235-016-0031-6.

Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist*, *49*(2), 59-69. doi:10.1080/00461520.2014.921572.

Burrage, M. (2008). "That's an interesting finding, but:" Postsecondary students' interpretations of research findings (Doctoral dissertation). Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/61578/mburrage_1.pdf?sequence=1&isAllowed=y

Carnevale, A. Strohl, J., & Smith, N. (2009). Help Wanted: Postsecondary education and training required. *New Directions for Community Colleges, 146*: 21–31. doi: 10.1002/cc.363

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (pp. 161–238). Mahwah, NJ: Erlbaum. Retrieved from: http://alumni.media.mit.edu/~bsmith/courses/mas964/readings/Glaser.pdf

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182. doi: 10.1016/0364-0213(89)90002-5

Crowell, A., & Schunn, C. (2016). Unpacking the relationship between science education and applied scientific literacy. *Research in Science Education, 46*(1), 129-140. doi: 10.1007/s11165-015-9462-1

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206–214. doi: 10.1016/j.learninstruc.2011.11.001

Durant, J. R. (1993). What is scientific literacy? In J. R. Durant & J. Gregory (Eds.), *Science and culture in Europe* (pp. 129–137). London: Science Museum.

Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77-84. doi: 10.1016/j.learninstruc.2011.11.001

Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science*, *7*(10), 454-469. doi: 10.1016/j.tics.2003.08.012.

Evans, J. & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*(4), 382-389. doi: 10.1080/13546780542000005.

Fong, G., Krantz, D., & Nisbett, R. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*(3), 253-292. doi: 10.1016/0010-0285(86)90001-0

Gobert, J., & Clement, J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, *36*(1), 39-53. doi: 10.1002/(sici)1098-2736(199901)36:1<39::aid-tea4>3.0.co;2-i

Golle, J., Mast, F. W., & Lobmaier, J. S. (2014). Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition and Emotion, 28(*2), 298-310. doi: 10.1080/02699931.2013.817383

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17,* 612–634. doi: 10.1016/j.learninstruc.2007.09.008

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449-455. Retrieved from http://psycnet.apa.org/doi/10.1037/0003-066X.53.4.449.

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

Huber, C. R., & Kuncel, N. R. (2015). Does college teach critical thinking? A meta-analysis. *Review of Educational Research,* 1987, 1-38. doi: 10.3102/0034654315605917.

Ibrahim, A., Seifert, C. M., Adar, E., & Shah, P. (2016). Using graphs to debias misinformation (in preparation).

Jenkins, E. W. (1994). Scientific literacy. In T. Husen & T. N. Postlethwaite, (Eds.), *The international encyclopedia of education (Volume 9, 2nd ed.,* pp. 5345–5350). Oxford, UK: Pergamon Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.

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. doi: 10.1111/1467-8624.00232.

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., Ulvik, M. (2006). Science students’ critical examination of scientific information related to socio-scientific issues. *Science Education,* 90(4), 632-655. doi:10.1002/sce.20133

Kosonen, P., & Winne, P. H. (1995). Effects of teaching statistical laws on reasoning about everyday problems. *Journal of Educational Psychology, 87*(1), 33. doi: 10.1037/0022-0663.87.1.33.

Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotive, W. (1988). *Developmental psychology series*. *The development of scientific thinking skills*. San Diego, CA, US: Academic Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. doi: 10.1037/0033-2909.108.3.480

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, & R. Lerner (Series Eds.) & K. A. Renninger, & I. E. Sigel (Vol. Eds.*), Handbook of child psychology: Vol. 4. Child psychology in practice (6th ed.)*. New York: John Wiley and Sons. doi: 10.1002/9780470147658.chpsy0405.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106-131. doi: 10.1177/ 1529100612451018.

Miller, J. D. (1996). Scientific literacy for effective citizenship. *Science/Technology/Society as Reform in Science Education.* Albany, NY: SUNY Press.

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Science communication: Public acceptance of evolution. *Science, 313,* 765-766. doi: 10.1126/science.1126746

Moore, D. W. (June 16, 2005). Three in four Americans believe in paranormal. *Gallup Poll News Service.* Retrieved from http://www.gallup.com/poll/16915/three-four- americans-believe-paranormal.aspx.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.*

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. doi: 10.1037/1089-2680.2.2.175.

Norcross, J. C., Gerrity, D. M., & Hogan, E. M. (1993). Some outcomes and lessons from a cross-sectional evaluation of psychology undergraduates. *Teaching of Psychology, 20*(2), 93-96. doi: 10.1207/s15328023top2002_6

Norris, S. P., & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching, 31*(9), 947-967. doi: 10.1002/tea.3660310909.

Norris, S. P., Phillips, L. M., & Korpan, C. A. (2003). University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. *Public Understanding of Science*, *12*(2), 123-145. doi: 10.1177/09636625030122001

Picardi, C. A., & Masick, K. D. (2013). *Research methods: Designing and conducting research with a real-world focus.* Thousand Oaks, CA: SAGE Publications.

Reis, H. T., & Judd, C. M. (2000). *Handbook of research methods in social and personality psychology.* Cambridge, UK: Cambridge University Press.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23,* 90–108. doi: 10.1006/ceps.1997.0959

Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuro-science information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1432-1440. doi: 10.1037/a0036844.

Rodriguez, F., Ng, A., & Shah, P. (2016). Do college students notice errors in evidence when critically evaluating research findings? *Journal on Excellence in College Teaching, 27*(3), 63-78. Retrieved from http://www.hackyourmotivation.com/uploads/9/7/5/8/97588042/scientific_reasoning.pdf

Rodriguez, F., Rhodes, R. E., Miller, K., & Shah, P. (2016). Examining the influence of anecdotal stories and the interplay of individual differences on reasoning. *Thinking & Reasoning, 22*(3), 274-296. doi: 10.1080/13546783.2016.1139506.

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. doi: 10.1037/0022-0663.91.3.497.

Sagan, C. (1996a). Does truth matter? Science, pseudoscience, and civilization. *Skeptical Inquirer, 20,* 28-33. Retrieved from https://www.csicop.org/si/show/does_truth_matter_science_pseudoscience_and_civilization

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T. & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review, 39,* 37-63. doi: 10.1016/j.dr.2015.12.001

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology, 71,* 55-89. doi: 10.1016/j.cogpsych.2013.12.004

Shah, P., Michal, A., Ibrahim, A., Rhodes, R., & Rodriguez, F. (2017). What makes everyday scientific reasoning so challenging? *The psychology of learning and motivation, 66,* (251-299). Retrieved from: doi.org/10.1016/bs.plm.2016.11.006

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science*, *11*, 433–448. doi: 10.1111/j.1467-7687.2008.00689.x

Sinatra, G. M., Kienhues, D., & Hofer, B. (2014). Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist*, *49*(2), 123-138. doi: 10.1080/00461520.2014.916216.

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction, 31*(2), 130-150. doi: 10.1080/07370008.2013.769996.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction, 21,* 22–33. doi: 10.1016/j.learninstruc.2009.10.001

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12,* 39–60. doi: 10.1016/s0959-4752(01)00015-9

Tal, A., & Wansink, B. (2016). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science, 25*(1), 117-125. doi: 10.1177/0963662514549688.

Trefil, J. (2008). Science education for everyone: Why and what? *Liberal Education*, *94*(2), 6-11. Retrieved from https://www.aacu.org/publications-research/periodicals/science-education-everyone-why-and-what

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*(2), 155-174.

doi: 10.1007/s10648-010-9134-7

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15*(6), 307-311. doi: 10.1111/j.1467-8721.2006.00458.x

Whoriskey, P. (4/3/2011). Requiring Algebra 2 in high school gains momentum. *The Washington Post.* Retrieved from https://www.washingtonpost.com/business/economy/requiring_algebra_ii_in_high_school_gains_momentum_nationwide/2011/04/01/AF7FBWXC_story.html?noredirect=on&utm_term=.a153d444a4bd

WorldHealth.net (6/12/2010). Sincere Smiling Promotes Longevity. *WorldHealth.net* Retrieved from: https://www.worldhealth.net/news/sincere-smiling-promotes-longevity/

Yeo, J., & Tan, S. C. (2010). Constructive use of authoritative sources in science meaning-making, *International Journal of Science Education*, *32*(13), 1739-1754. doi: 10.1080/09500690903199564

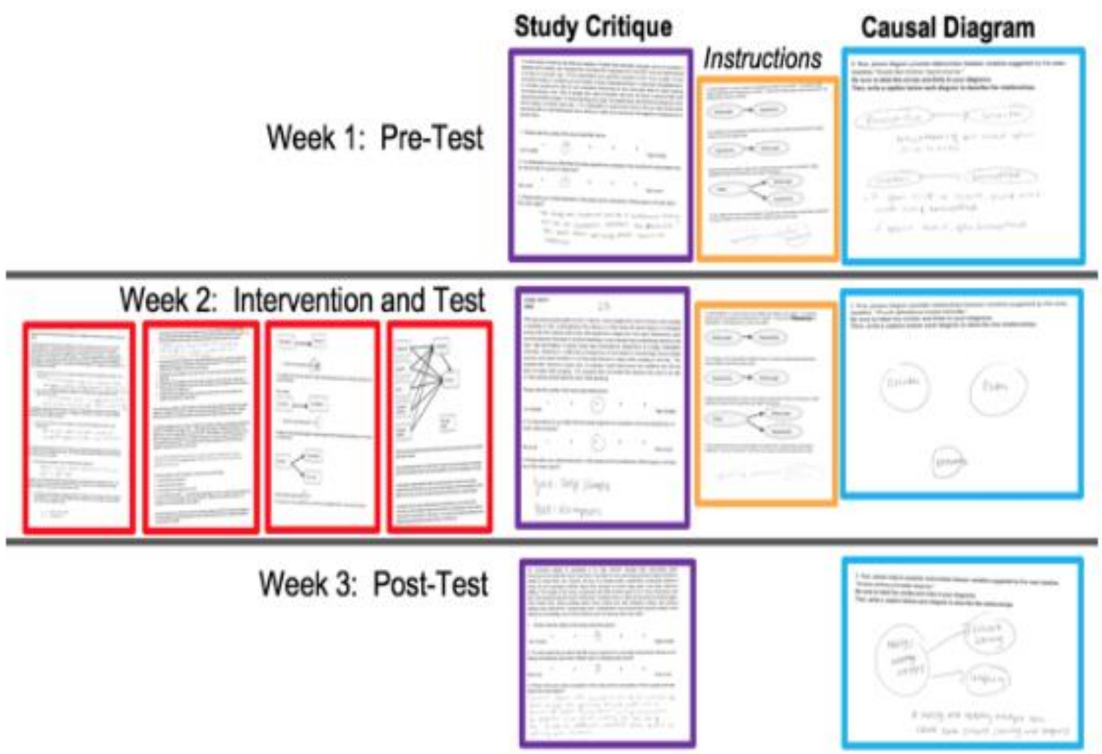


Figure 1. Graphic depiction of questionnaire content for each of the three sessions.

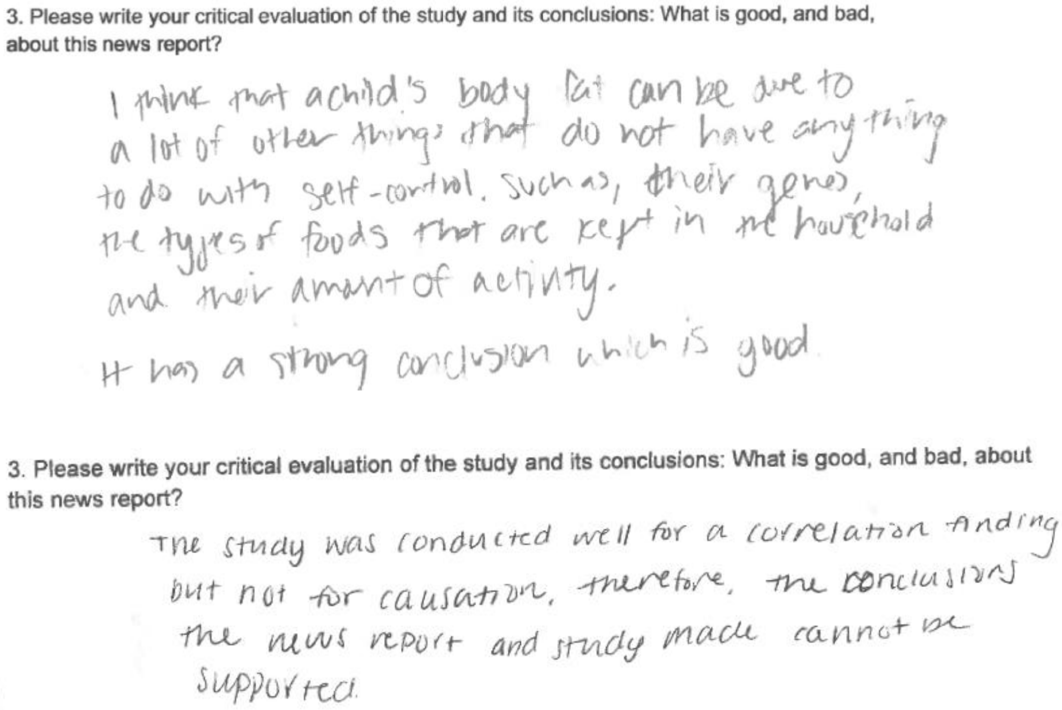
Please rate the quality of the study described above.

	1	2	3	4	5	
Low Quality						High Quality

2. To what extent do you think that the study supports the conclusion that one should listen to music while studying?

	1	2	3	4	5	
Not at all						Very much

Figure 2. Example of a media story with a causal claim and the two ratings (quality and support) questions.



3. Please write your critical evaluation of the study and its conclusions: What is good, and bad, about this news report?

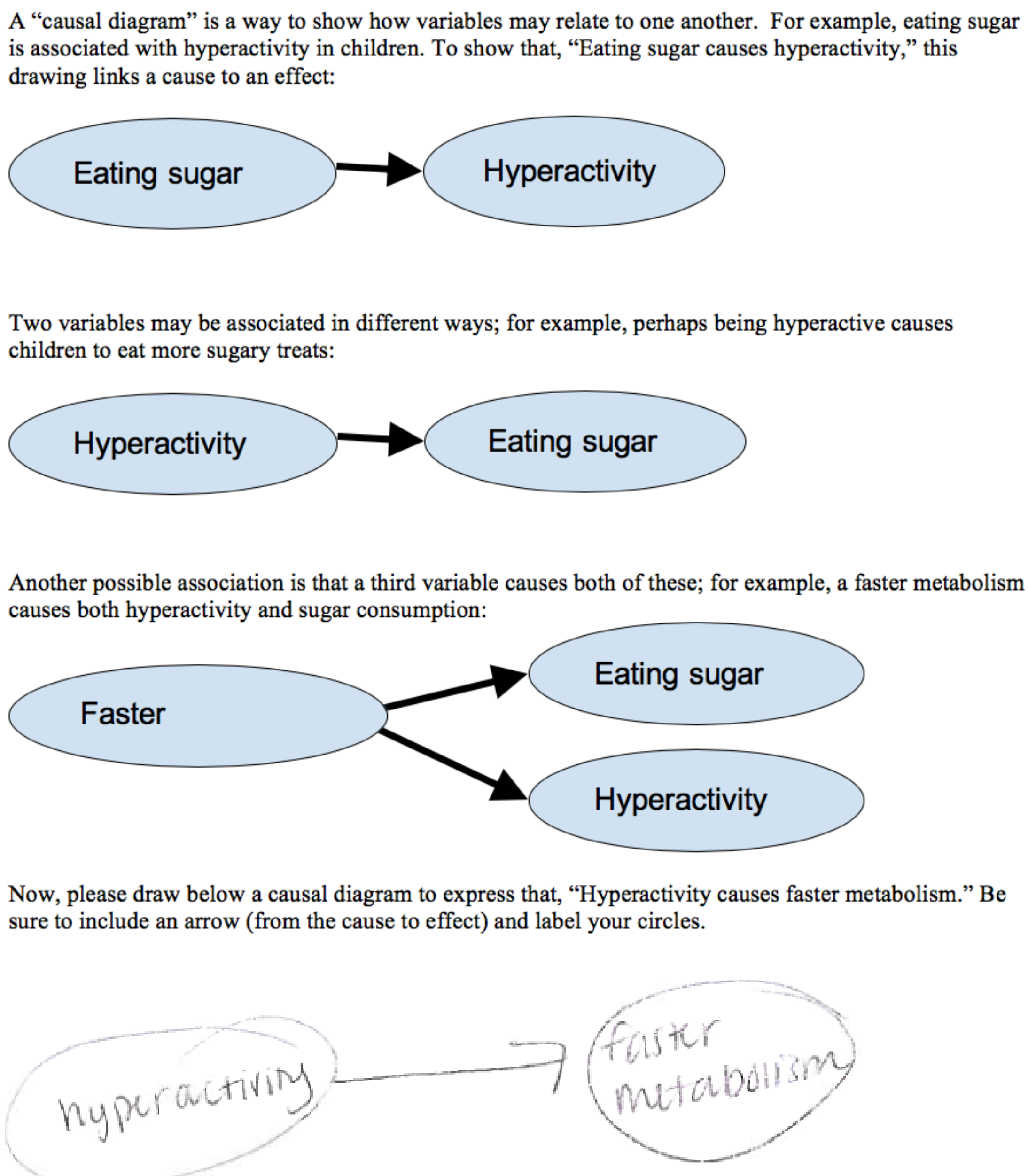
I think that a child's body fat can be due to a lot of other things that do not have anything to do with self-control. Such as, their genes, the types of foods that are kept in the household and their amount of activity.

It has a strong conclusion which is good.

3. Please write your critical evaluation of the study and its conclusions: What is good, and bad, about this news report?

The study was conducted well for a correlation finding but not for causation, therefore, the conclusions the news report and study made cannot be supported.

Figure 3. Two examples of students' critical evaluation responses; the top panel response suggests alternative causal factors not mentioned in the study, while the bottom emphasizes that a correlational finding is not indicative of a causal relationship.



A "causal diagram" is a way to show how variables may relate to one another. For example, eating sugar is associated with hyperactivity in children. To show that, "Eating sugar causes hyperactivity," this drawing links a cause to an effect:

Two variables may be associated in different ways; for example, perhaps being hyperactive causes children to eat more sugary treats:

Another possible association is that a third variable causes both of these; for example, a faster metabolism causes both hyperactivity and sugar consumption:

Now, please draw below a causal diagram to express that, "Hyperactivity causes faster metabolism." Be sure to include an arrow (from the cause to effect) and label your circles.

Figure 4. Instructions for causal diagrams included several example diagrams. The correct response from a student's booklet is shown at the bottom.

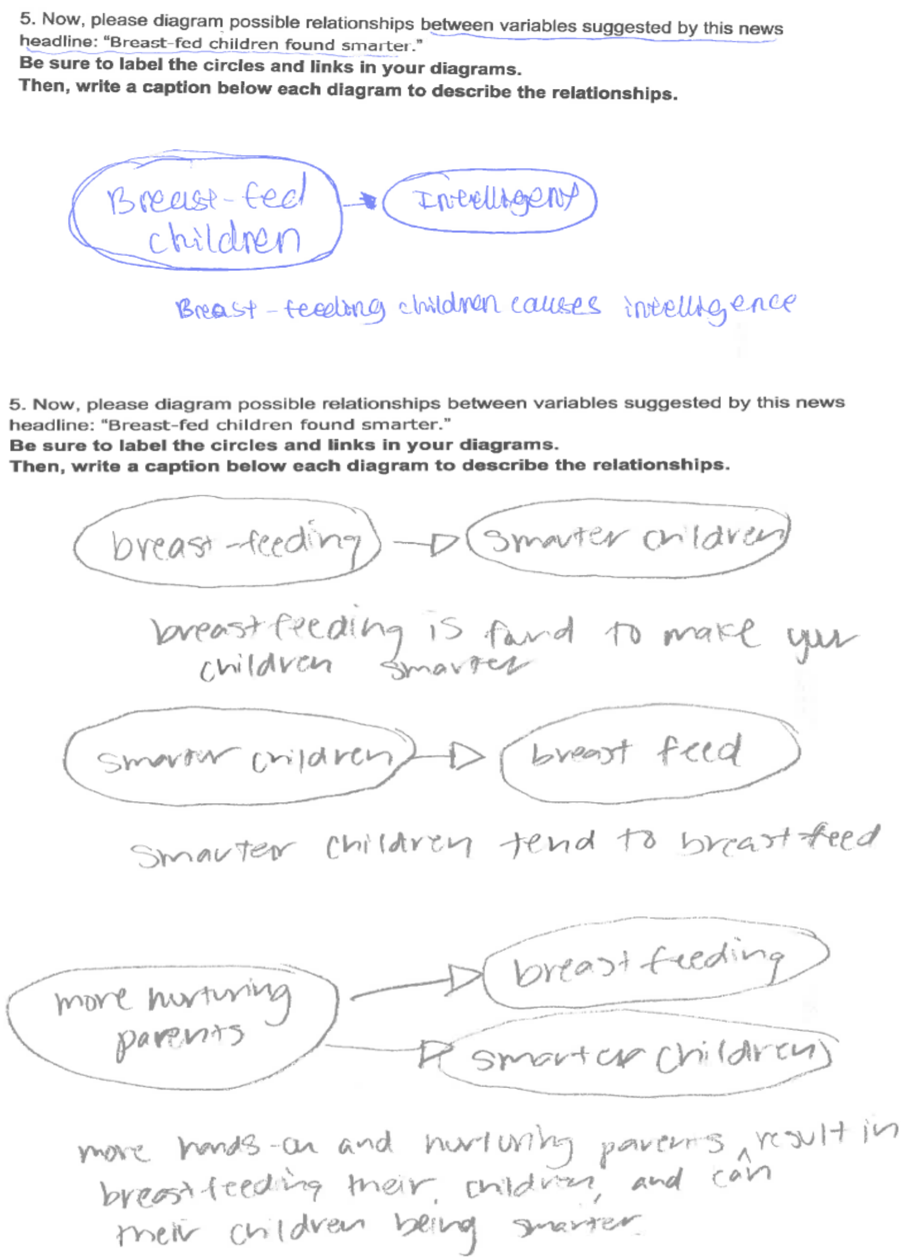


Figure 5. Two examples of student responses to the headline Causal Diagram task. The student's response in the top panel depicts one causal relationship between the two variables, while the student's response in the bottom panel includes three diagrams reflecting three different causal relationships.

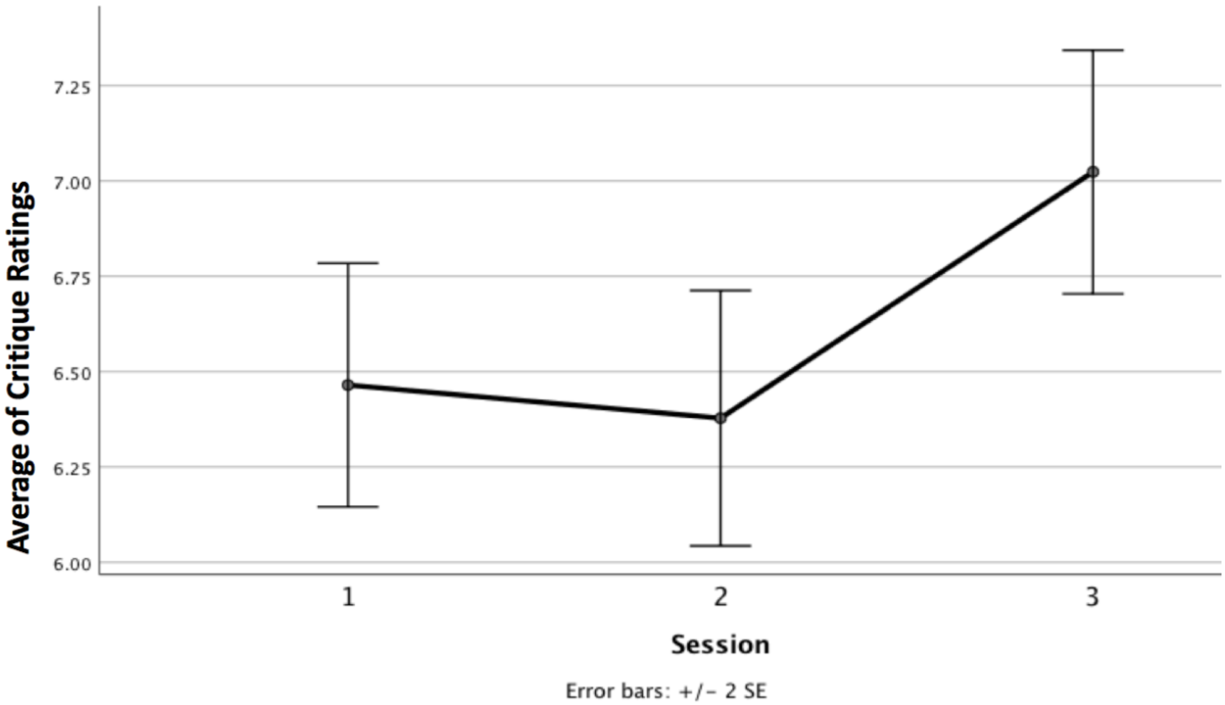


Figure 6. Average combined quality and support ratings (ranging from 2 to 12) across sessions.

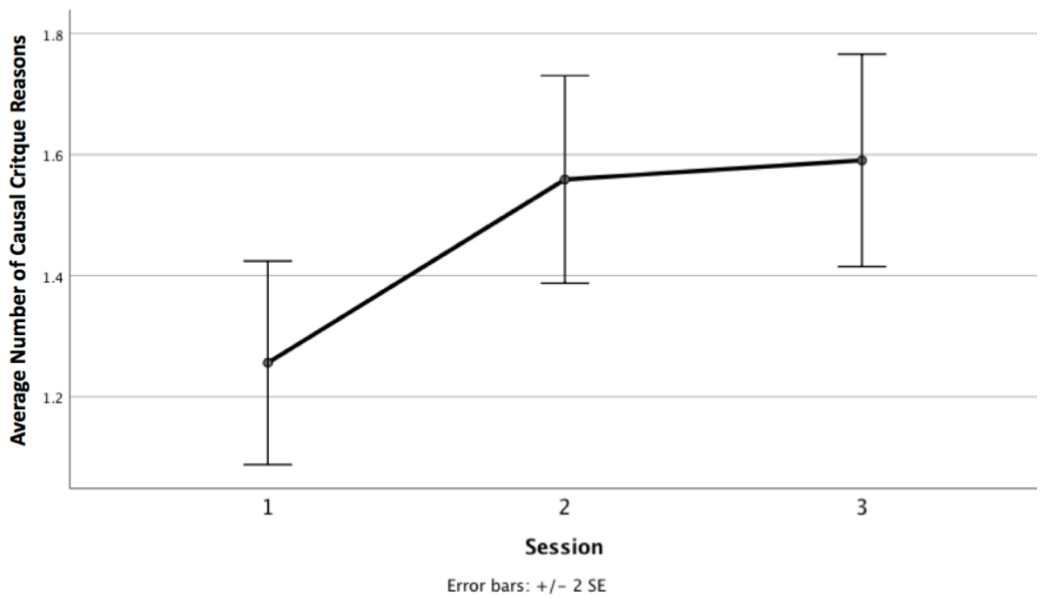


Figure 7. Average number of causal reasons given in open-ended study critiques across sessions.

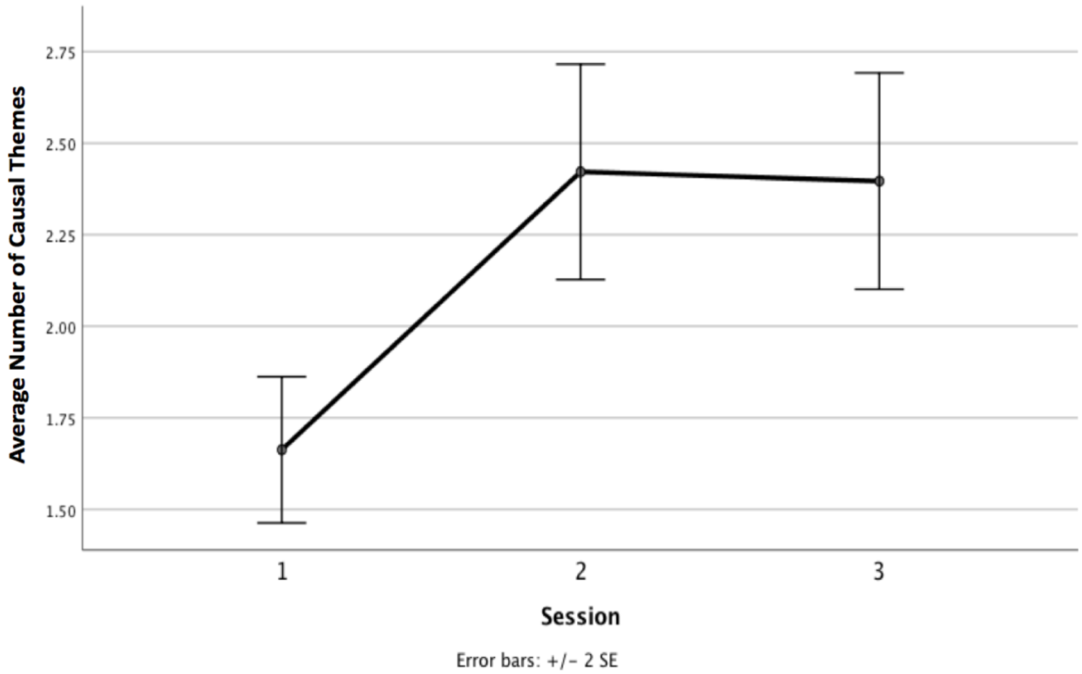


Figure 8. Average number of alternative causal themes in subjects' diagrams across sessions.

Appendix A

Intervention in Session 2

Please read the following packet and answer the questions to the best of your ability as you go along.

In 2006, researchers at Educational Testing Service (ETS) conducted an experiment in which they followed students for 12 years starting in 8th grade. They found that 84% of the top-tier workers (receiving the highest pay) had taken Algebra II (or higher) classes in high school. In contrast, only 50% of workers in the lowest-tier had taken Algebra II. This result suggests that requiring students to take Algebra II would benefit all students and can prepare them for better jobs in the new Millennium. Twenty states, including Michigan, have passed laws making Algebra II a graduation requirement for all of their high school students.

- Does the ETS study convince you that taking Algebra II would be beneficial for students? Why or why not?
- Do you think that requiring Algebra II based on the ETS study was a good decision for the state? Why or why not?

The Algebra II requirement has now fallen out of favor in some states, and they are now reversing or watering down their Algebra II requirement. In Michigan, students can now take a "Career Technical Education" course instead of Algebra II.

- How do you think that no required Algebra II course will affect Michigan students' career opportunities?

Some people (and legislators) view the ETS study as proving that taking Algebra II causes students to have better chances of getting top-tier jobs. That certainly seems plausible because *having math skills should help you get a good job!*

- Why might taking Algebra II lead to students getting a better job?

However, it is *not necessarily true* that taking Algebra II would lead directly to getting better jobs. It is possible that the connection between Algebra II and higher job status is that the students who took Algebra II differed in other ways from people who did not.

- In what ways might students in Algebra II differ from students who don't take it? Who chooses to take Algebra II? Try to think of other differences that might lead some students towards top-tier jobs. List at least 2 different things.

1.
2.

- Think back to the ETS study, when students had to choose whether to take Algebra 2. Why would a high school student decide to take it, and why would a student decide not to take it? Try to think of at least one new reason a student would take it, and one reason not to take it.

1.

- Think about the reasons listed below, and judge whether each reason might also explain why algebra students end up in better jobs. Mark each sentence with a "T" for true and and "F" for false based on whether you think it is a good reason.

1. Students who chose Algebra II were also *smarter*, so they did well in school and ended up in better jobs.
2. Students who chose Algebra II were also *going to college*, so they ended up in better jobs.

3. Students who chose Algebra II were from *richer* families who would help them with connections so they ended up in better jobs.

4. Students who chose Algebra II went to *better high schools* (with more math classes), and therefore they ended up in better jobs.

From the previous questions, notice that there are other reasons that students might both have taken Algebra II and gotten a top-tier job. It might *look like* what the students learned in Algebra II helped them get good jobs, but it could have been one of these other reasons that was the real cause.

The Michigan legislators seem to make a mistake in their decision making (as all people sometimes do) called the "correlation-to-causation" error. Just because two things are related (like Algebra II and better jobs), you can't decide that one *causes* the other. Taking Algebra II and getting a good job may both be caused by something else; for example, being good in school or being *in* a good school, or having parents that are doctors. This is called a *third variable* explanation, where two things "go together" because of some other (third) cause.

Take a moment to pause and think: Most of us already know that "correlation does not imply causation;" but many times, especially when the cause makes sense, we forget to question our assumptions about causes.

If any two variables A and B are related, it can mean one of several things:

1. It could mean that A causes B.
2. It could mean that B causes A.
3. It could mean that C causes *both* A and B.

Or, A and B both can cause C. It could be really complicated, in which A causes B which causes more A, and so on. Or, that A causes B, but C also impacts B. Whenever you evaluate evidence, it is important to think through these various options.

Appendix B

Study Critique and Causal Diagram Tasks

COMPUTERS-GLASSES
A recent study funded by the National Institute of Health links extended computer use to an increase in glasses and contacts use. Researchers recruited 600 employees from the NYC area and administered a survey on computer use. Of the respondents who used the computer for 30+ hours a week, ⅔ wore corrective lenses or contacts and had severe myopia (nearsightedness) or hyperopia (farsightedness). In contrast, people who did not use computers extensively at work were less likely to report wearing corrective lenses: Only 10% of people who used computers less than 30 hours a week at their jobs required corrective lenses. To avoid harming your eyes, the researchers recommend avoiding too many hours using a computer each day. If it is impossible to avoid screen time on the job, they recommend speaking with an ophthalmologist about what you might do to counteract the negative consequences of screen time.

SILENCE OR MUSIC
Although some people prefer to work in silence, many people opt to listen to music while working or studying. In fact, a brief glimpse into a library or coffee shop will reveal dozens of individuals poring over their laptops and books with earphones wedged into their ears. Researchers have recently become interested in whether listening to music actually helps students pay attention and learn new information. A recent study was conducted by researchers at a large midwestern university. Students (n = 450) were surveyed prior to final exams in several large, lecture-based courses, and asked whether or not they had listened to music while studying for the final. The students who listened to music had, on average, higher test scores than students who did not listen to music while studying. The research team concluded that students who want to do well on their exams should listen to music while studying.

PARENT-SELF CONTROL
An important aspect of parenting is to help children develop their self-control skills. Developmental scientists have long been interested in how parenting practices impact children's ability to make their own choices. 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 400 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. Researchers recommend that parents should avoid being too controlling, and let their children learn to develop their own skills.

CHURCH-HEALTH
Now, please diagram possible relationships between variables suggested by this news headline: "Church attendance boosts immunity." Be sure to label the circles and links in your diagrams. Then, write a caption below each diagram to describe the relationships.

SMILING-LONGEVITY

Now, please diagram possible relationships between variables suggested by this news headline: “Sincere smiling promotes longevity.” Be sure to label the circles and links in your diagrams. Then, write a caption below each diagram to describe the relationships.

CHURCH ATTENDANCE–IMMUNITY

Now, please diagram possible relationships between variables suggested by this news headline: “Church attendance boosts immunity.” Be sure to label the circles and links in your diagrams. Then, write a caption below each diagram to describe the relationships.