The Influence of Consequence Value and Text Difficulty on Affect, Attention, and Learning while Reading Instructional Texts

Caitlin Mills\textsuperscript{a}, Sidney K. D’Mello\textsuperscript{a,b}, and Kristopher Kopp\textsuperscript{a}

Departments of Psychology\textsuperscript{a} and Computer Science\textsuperscript{b}, University of Notre Dame, IN, 46556, U.S.A., cmills4@nd.edu, sdmello@nd.edu, kkopp@nd.edu

Corresponding Author
Caitlin Mills
118 Haggar Hall
University of Notre Dame
Notre Dame, IN 46556
USA
cmills4@nd.edu
Highlights

- Participants studied four instructional texts on research methods
- Consequence value and text difficulty were manipulated in a 2×2 design
- Consequence value had direct effects on affect and knowledge transfer
- Text difficulty had direct effects on mind wandering and reading time
- Arousal mediated the relationship between consequence value and knowledge transfer
- Text difficulty moderated the mediation through arousal
Abstract

The present study investigated how consequence value influences affect, attention, and learning while reading instructional texts, and if text difficulty moderates these effects. Participants studied four instructional texts on research methods in a 2 × 2 consequence value (high vs. low) × text difficulty (easy vs. difficult) within-subjects experiment. Consequence value was manipulated by assigning two of the four texts as having high value and the other two as having low value with respect to a performance goal on a subsequent test, while text difficulty was manipulated via experimenter-created easy and difficult versions of the texts. We hypothesized that consequence value would induce mild anxiety, which would focus attention and facilitate learning, and that text difficulty would moderate the influence of consequence value. Partially consistent with the predictions, high consequence value led to lower valence, higher arousal, longer reading times, and positively predicted knowledge transfer. Arousal mediated the relationship between consequence value and knowledge transfer, but only when the texts were difficult, thereby suggesting moderated mediation.

Keywords: affect, attention, learning, motivation, consequence value
The Influence of Consequence Value and Text Difficulty on Affect, Attention, and Learning while Reading Instructional Texts

1. Introduction

Educators are often faced with the challenge of motivating students to learn. This can be particularly difficult, especially when students are unmotivated and uninterested in the subject itself. In these cases, extrinsically-valued rewards are a commonly-used motivational strategy. Extrinsic value refers to students’ perceptions of the value of a learning activity as it relates to goals unrelated to the activity itself (Pekrun, Goetz, Titz, & Perry, 2002). These goals can be in the form of receiving a reward as well as avoiding a consequence.

The effects of extrinsic value have received considerable attention, particularly with respect to the influence of rewards on task performance and intrinsic motivation (Deci & Ryan, 1985; Eisenberger, Pierce, & Cameron, 1999; Fryer, 2011; Pierce, Cameron, Banko, & So, 2012). Meta-reviews on the relationship between rewards and task performance have reported mixed findings, including positive, negative, and null effects (Deci, Koestner, & Ryan, 2001; Pierce et al., 2012). For example, monetary incentives have been both positively and negatively related to performance (Bettinger, 2012; Fryer, 2011). The effects of extrinsic rewards on intrinsic motivation have also been a topic of debate. Self-determination theory posits that rewards decrease intrinsic motivation because they undermine students’ sense of autonomy (Deci, Koestner, & Ryan, 1999; Deci et al., 2001; Deci & Ryan, 2012), a claim that has received empirical support (see Deci, Koestner, & Ryan, 1999 for a review). However, conflicting evidence suggests that rewards can also have positive effects on intrinsic motivation (Cameron, Pierce, Banko, & Gear, 2005; Hagger & Chatzisarantis, 2011; Pierce et al., 2012). It might be the
case that rewards can be effective in appropriate situations, such as when the material is uninteresting or the reward is unexpected (Cordova & Lepper, 1996; Sansone, 2000).

In many real-world situations, negative consequences accompany extrinsic rewards upon failure to reach a goal (e.g., reward of getting a passing grade vs. consequences of failing, such as having to attend summer school). Self-determination theory (SDT) predicts that negative consequences should also lower intrinsic motivation (Roth, Assor, Niemiec, Ryan, & Deci, 2009). Research in support of this hypothesis has indicated that students perceive lower autonomy and intrinsic motivation when they feel pressure to perform an activity in order to avoid a negative consequence (Assor, Roth, & Deci, 2004; Hancock, 2001; Roth et al., 2009). However, contrasting evidence also suggests that negative consequences can be linked to increased motivation for achievement as well as performance during learning (Miller, Greene, Montalvo, Ravindran, & Nichols, 1996; Smith & Smith, 2002). It might be the case that, similar to extrinsic rewards, negative consequences can also be effective in specific situations, although existing research does little to reconcile the conflicting findings.

Previous research on the effects of rewards and negative consequences1 has focused on motivation and performance as outcome variables (Bettinger, 2012; Cameron et al., 2005; Hagger & Chatzisarantis, 2011). Affect and attention are equally important, yet often overlooked, processes to consider as they have been shown to play critical roles during learning (Pekrun & Linnenbrink-Garcia, 2014; Sweller, 1988). Affect and attention are conceptually distinct from motivation, as motivation is more closely related to students’ goals prior to a learning activity, while affect and attention occur during a learning activity and are influenced by a multitude of factors beyond motivation (Anderman & Patrick, 2012). Moreover, models of motivation do not

---

1 Henceforth, negative consequences are referred to simply as consequences. Positive outcomes are called rewards.
typically place a primary focus on relatively short-term states, such as affect and attention (Deci & Ryan, 2012; Eccles & Wigfield, 2002). Thus, a focus on affect and attention might offer an alternative explanation for how consequence value influences learning. We consider this possibility in the present research.

The primary goal of the current experiment was to investigate the influence of negative consequence value on affect, attention, and learning and to assess if difficulty moderated this effect. Students read four instructional texts on scientific research methods. Affect was measured via self-reports of valence and arousal at multiple points during reading. Attention was primarily measured via periodic self-reports of mind wandering and secondarily via overall reading time. Learning was measured with knowledge tests after the learning session. Negative consequence value (referred to as consequence value) was manipulated by assigning two of the four texts as having high consequence value and two as having low consequence value with respect to a performance goal on a subsequent test. Failure to meet the performance goal resulted in an undesirable penalty (having to read more texts). This manipulation attempted to mirror a real-world situation where, in addition to potentially being rewarded for achievement (e.g., free time on the computer), failure itself has negative consequences (e.g., having to re-take a test). Text difficulty was manipulated via experimenter-created easy and difficult versions of the texts.

We test two specific hypotheses grounded in the control-value theory (CVT) of academic emotions. CVT posits that the affective states that arise during learning are based on appraisals of subjective control and subjective value (Pekrun & Stephens, 2010; Pekrun, 2006). Subjective control refers to students’ appraisals of their own abilities to take action in order to achieve a desirable outcome (Pekrun & Stephens, 2010). When a task is difficult relative to a student’s ability, the student will perceive having lower subjective control towards achieving the desired
outcome, and vice versa for easy tasks (Pekrun, 2006). Subjective value pertains to the perceived interest/importance of the activity and can be either negative or positive, depending on students’ object of focus (Pekrun, 2006), which can be the activity itself (e.g., enjoyment or dislike) or the outcome (e.g., success or failure). Positive subjective value results when the focus is on enjoyment of the activity or the reward of a successful outcome. Alternatively, negative subjective value occurs when the focus is on negative feelings associated with the activity or if the consequences of failure are aversive.

Consistent with circumplex models of affect (Pekrun & Linnenbrink-Garcia, 2012; Russell, 2009), CVT posits that both valence (i.e., positive and negative feelings) and arousal (i.e., level of activation) are the basic components of affect. As depicted in Table 1, discrete affective states arise from different combinations of valence and arousal. In particular, positive subjective value, when combined with low arousal, is associated with a state of relaxation or calmness, but is more akin to state of engagement (Pekrun & Linnenbrink-Garcia, 2012; Pekrun et al., 2002) when paired with high arousal. Negative subjective value can lead to boredom and disengagement when combined with low levels of arousal. However, negative subjective value can lead to mild anxiety when arousal is moderate. Importantly, this form of mild anxiety is expected to be facilitative rather than harmful to learning, as elaborated in Hypothesis 1 (the consequence value hypothesis).

The consequence value hypothesis is multicomponential in that it includes affect, attention, and learning. The first component is a derivation from CVT in that the consequence value manipulation will be subjectively appraised as being negative due to undesirable consequences of failure. Focus on the threat of failure will increase arousal, thereby resulting in a
state of anxiety (Pekrun et al., 2002), which should be reflected in our measures via lower valence and higher arousal for the high vs. low consequence value texts.

It is widely known that anxiety focuses attention in order to identify, avoid, or eliminate potential threats in the environment (Fielder & Beier, 2014). In addition to influencing affect, consequence value is also expected to influence attention, measured in the present study via attentional lapses (or mind wandering) and reading time. Mind wandering is an unintentional shift of attention from task-related thoughts to task-unrelated thoughts, and has been linked to negative performance outcomes (Randall, Oswald, & Beier, 2014; Smallwood & Schooler, 2015). We predict mind wandering will occur less frequently during high consequence value texts due to attentional focus associated with anxiety. We also predict that consequence value will influence reading time, which is taken as an indicator of attention and effort (Guthrie et al., 2012; Ponitz et al., 2009). Participants should spend more time reading the high consequence value texts because of increased pressure to comprehend those texts.

Finally, with respect to learning, it is hypothesized that high consequence value texts should facilitate learning compared to low value texts. This is because the anxiety triggered by the threat of a negative outcome can motivate performance (in part via focused attention) due to a desire to avoid the negative outcome (Martin & Marsh, 2003; Pekrun et al., 2002; Smith & Smith, 2002).

In summary, the consequence value hypothesis posits that consequence value will be subjectively appraised as negative and will be associated with an affective state akin to mild anxiety (lower valence and higher arousal), more attention (less mind wandering and longer reading times), and increased learning for the high vs. low value texts. Since affect and attention are expected to influence learning, we also posit that the effect of consequence value on learning
will be mediated through affect and attention. This prediction is based on the notion that increased levels of arousal should positively relate to learning since arousal is hypothesized to signal information as important for encoding and later retrieval (Storbeck & Clore, 2008) and due to the well-known relationships between arousal and performance (Strain, Azevedo, & D’Mello, 2013; Yerkes & Dodson, 1908). Mind wandering is expected to share a negative relationship with learning since mind wandering disrupts the creation of a mental model from the text as attention is directed toward unrelated off-task thoughts (Smallwood, Fishman, & Schooler, 2007; Smallwood & Schooler, 2006). Reading time should positively correlate with learning because we assume that reading time is indicative of effort and persistence (Guthrie et al., 2012). We did not make a prediction for valence since both positive and negative valence can facilitate and hinder learning. For example, previous research suggests that negative feelings are associated with more concrete representations and better recall (Beukeboom & Semin, 2006; Storbeck & Clore, 2005), while positive feelings have been related to problem solving and inference generation (Isen, Daubman, & Nowicki, 1987). The hypothesized mediation paths are presented in Figure 1.

It is important to note that the aforementioned links between anxiety, attention, and learning are only assumed to apply when the level of anxiety is mild to moderate. For example, both attention and learning are likely to be hindered when anxiety is too high, partly because anxious thoughts consume working memory resources (Beilock & Carr, 2005; Eysenck, Derakshan, Santos, & Calvo, 2007). This raises the question of what circumstances constitute mild productive anxiety? In line with this, our second hypothesis predicts that difficulty will moderate the effect of consequence value on affect, attention, and learning. Specifically, the effects of consequence value may only apply under circumstances when task demands also
CONSEQUENCE VALUE AND DIFFICULTY

promote mild anxiety. For example, if students perceive task demands as being too difficult, students may feel pressure to the point where unproductive anxiety sets in. Conversely, if the task demands are too easy, students may feel confident in their ability to learn the material and not feel the same sense of anxiety or pressure to avoid failure.

If students perceive the task as being moderately difficult, they may feel a productive level of mild anxiety because of the challenge, which is related to positive performance outcomes (Yerkes & Dodson, 1908). In other words, when consequence value is high (e.g., important to future goal) and the task is of moderate difficulty, students may feel productive levels of pressure and anxiety, which can decrease mind wandering (due to increased attentional focus on the material), increase effort, and facilitate learning (Martin & Marsh, 2003; Pekrun et al., 2002). Thus, Hypothesis 2, or the difficulty moderation hypothesis, predicts that the direct effects of consequence value will be moderated by task demands, such that productive states (mild anxiety in this case), less mind wandering, longer reading times, and increased learning will only occur when task demands are moderately difficult compared to being very easy.

To summarize, the relationship between consequence value and learning is expected to be mediated through mild anxiety and increased attention (consequence value hypothesis). Task demands are predicted to moderate the direct effects of consequence value (difficulty moderation hypothesis). Taken together, it is possible that the mechanism by which consequence value influences learning through affect and attention may be different at varying levels of task demands (known as moderated mediation). Moderated mediation attempts to explain both how and when an effect occurs (Bauer, Preacher, & Gil, 2006; Preacher, Rucker, & Hayes, 2007). When task demands are difficult, increased levels of productive anxiety and attention may play a more important mediating role in comparison to tasks with low demands, possibly due to the
increased pressure. Thus, the *difficulty moderation hypothesis* further predicts that the effect of consequence value will show stronger evidence of mediation through affect and attention when task demands are high versus low.

The *difficulty moderation hypothesis* was tested by manipulating text difficulty (easy vs. difficult) in addition to consequence value. Thus, each text was either *easy* or moderately *difficult* to read, and with either low or high consequence value.

2. Method

2.1 Participants

Participants were 187 undergraduate students from a large public U.S. university and a medium-sized private U.S. university (N = 105 and N = 82) who participated for course credit. The average age was 19.7 years (SD = 2.69, 58% female). The majority of the participants were Caucasian (54.3%), with 28.7% being African American, 6.9% Asian, 6.4% Hispanic, and 3.7% Other.

2.2 Instructional Texts

The texts were adapted from texts that accompany the educational game *Operation ARA!* (Millis et al., 2011). Each text focused on one of four research methods topics: replication, experimenter bias, making causal claims, and dependent variables. Each text had the following components: (1) a title, (2) a case study that consisted of an example of a real world application of the topic, and (3) explanations and additional examples for three core concepts relevant to the topic. For example, the text on experimenter bias used the story of Clever Hans as the main case study to demonstrate the concept of bias. The three specific concepts covered were blind/ double-blind experiments, the ways bias might occur, and the ways participants are affected by bias. Each text was approximately 1,500 words in length.
2.3 Design and Manipulations

This experiment used a 2 (consequence value: low vs. high) × 2 (difficulty level: easy vs. difficult) within-subjects design in which participants read four texts on scientific research methods topics. We chose to use a within-subjects design in order to minimize the influence of individual differences on the dependent measures. The study discussed in the current paper was independent and not part of a larger investigation. The data collection process was, however, used as an opportunity to pilot a few additional individual difference measures that are not germane to the current study and are therefore not analyzed or discussed.

The order of the four topics, conditions, and assignment of condition to topic was counterbalanced across participants using a Graeco-Latin Square design. There were four conditions in the 2 × 2 design: high value – difficult, high value – easy, low value – difficult, low value – easy. The four conditions were paired with the four topics to form a 4 × 4 Graeco Latin square, such that every condition and topic occurred precisely once in each row and once in each column. We then created a second Graeco Latin square in order to include four additional orderings of condition, a practice recommended to reduce error when variability is unknown (Grant, 1948; Mandl, 1985). Thus, using two squares helped assuage the concern of order effects across the four conditions. Finally, a total of eight orderings (2 squares) were repeated twice to counterbalance test order (AB versus BA). A total of 16 participants were needed to complete a full round of the design.

It is important to clarify what is meant by low and high consequence value and easy and difficult texts before elaborating the experimental manipulations in detail. Low consequence does not mean zero consequence. It means some consequence does exist, but that there is more for the high consequence value texts. In the complete absence of value, CVT would predict an
overall lack of motivation to learn. Similarly, difficult does not mean unreadable. Instead, the difficult texts were still understandable but were expected to have higher task demands and require more effort to read compared to the easy texts.

2.3.1. Consequence Value Manipulation. The consequence value manipulation was implemented by first incentivizing participants to do well on the reading task by informing participants they would have to stay longer to read additional text if they did not meet a minimum threshold on a subsequent knowledge test. This also meant that they could leave the study early if they scored high enough; they were informed of this fact. Consequence value was manipulated by informing participants that questions on two of the four texts (i.e., the high consequence texts) would count three times more towards their test score than questions on the remaining two texts (i.e., low consequence value texts). Participants were informed of the weightings (3 or 1) before reading each text.

2.3.2. Text Difficulty Manipulation. Easy and difficult text versions were created for each of the four topics, resulting in eight texts in all. Participants received easy versions of the texts for two topics and difficult versions for the remaining two topics. Difficulty was manipulated by modifying the original texts based on specific dimensions that have been identified as playing an important role in text readability: verb cohesion and syntactic simplicity (Graesser, McNamara, & Kulikowich, 2011). Verb Cohesion is based on the degree to which there are overlapping verbs in a text, where overlapping verbs make the text more cohesive and narrative. Syntactic Simplicity is based on the simplicity of the syntactic structures in the text. Thus, easy versions of the text consisted of shorter, simpler sentences with higher frequency words (e.g., replacing the low-frequency word “assemblage” with a higher-frequency word like “group”). The easy versions contained more overlapping verbs, such that similar verbs and phrases represented
CONSEQUENCE VALUE AND DIFFICULTY

recurring ideas as opposed to a novel phrase at every occurrence (e.g. the verb “cause”
repeatedly used to represent the effect of an independent variable). They were also more
narrative in style, and contained more nouns as opposed to ambiguous pronouns. Difficult
versions had more complex, longer sentences with lower frequency words (see Table 2 examples
of the text difficulty manipulations). Different verbs and phrasings were also used to represent
similar ideas, so that readers would encounter a novel phrasing for similar ideas and less overlap.
For example, the verbiage “caused” could be encountered first, followed by “had an effect on”
and “influenced,” all of which represented the same idea.

Importantly, text length (mean of 1492 and 1503 words for easy and difficult texts,
respectively) and content were kept consistent across easy and difficult versions. In order to
verify the text difficulty manipulation, we performed an analysis of each text using Coh-Metrix,
a computational tool for multidimensional textual analysis (Graesser, McNamara, Louwerse, &
Cai, 2004). We focused the two manipulated linguistic features (verb cohesion and syntactic
simplicity) that have been linked to text difficulty (Graesser et al., 2011). Table 3 presents
descriptive statistics for both features. Higher values for the Coh-Metrix features indicate that the
text is easier to read and understand.

Differences were expected in the Coh-Metrix features, since they were targeted during
the text manipulation process. Therefore, additional analyses were performed using three
readability metrics to assess the linguistic features of the easy and difficult texts: Flesch-Kincaid
Grade Level (FKGL; Klare, 1974), Automated Readability Index (Kincaid, Fishburne Jr, Rogers,
& Chissom, 1975), and Pearson Reading Maturity Metric (Nelson, Perfetti, Liben, & Liben,
2012). FKGL and ARI are computed based on surface-level features, like the number of
sentences, words, and syllables. Pearson Reading Maturity Metric is based on a combination of
surface-level features (e.g., number of words) and deep characteristics of the text (e.g., word meaning and frequency), calculated using a statistical model of language called Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998). All three of these metrics approximate the level of text difficulty in reference to the grade level of the analyzed text. For example, a score of 7 would indicate a 7th grade reading level. Paired samples t-tests were conducted between the easy and difficult texts for each of the six textual features in Table 3. There were significant ($p < .05$) differences between the easy and difficult texts for all six features, thereby indicating the success of the manipulations.

2.4. Measures

2.4.1. Topic Interest (Individual Difference Covariate). Interest in the overall topic of research methods was expected to correlate with our dependent variables (affect, attention, and learning). To statistically co-vary out the effect of topic interest, we adapted a topic interest measure from Linnenbrink-Garcia et al. (2010) as a measure of participants' baseline interest in learning about research methods prior to engaging in the learning task. Sample questions included: “I am not at all interested in learning about research methods,” and “I think it is important to know what research methods are all about.” Responses ranged from 1 (strongly disagree) to 6 (strongly agree).

2.4.2. Manipulation Checks. Two questions about student perceptions of difficulty and consequence value were presented after each text was read. Perceived difficulty was measured with the following two questions: (1) “How difficult did you find this text?” and (2) “How well do you think you understood this text?” Participants responded by selecting a point on a 6-point scale between (1) “very easy” and “very difficult” and (2) “very poorly” and “very well,” respectively. Perceived consequence value was measured with the following two questions that
assessed value and pressure: (1) “How important do you think this text was for the test after the readings?” and (2) “How much pressure did you feel while reading this text?” Similarly, responses were on a 6-point scale from (1) “very worthless” to “very useful” and (2) “very little” to “very much,” respectively. These four perception questions were analyzed as manipulation checks to assess if the experimental manipulations of consequence value and text difficulty influenced students’ perceptions of the texts in expected ways. For example, participants’ ratings of value should be higher after reading a high compared to low consequence value text.

2.4.3. Affect. Valence and arousal were measured using the Affect Grid developed by Russell, Weiss, & Mendelsohn (1989). The Affect Grid is theoretically grounded in the circumplex model of affect and has been validated by Russell and colleagues as a single-item affect measurement (see Russell et al., 1989, for Psychometric validation of the Affect Grid). It was presented as a 9 × 9 grid with two dimensions (valence × arousal). The arousal dimension ranges from (1) sleepiness to (9) high arousal. The valence dimension ranges from (1) unpleasant feelings to (9) pleasant feelings. Participants selected a cell on the Affect Grid to indicate their current affective state.

2.4.4. Mind Wandering. Mind wandering was measured by periodically probing participants during reading to indicate if they were thinking of task-unrelated thoughts. This is the most common method used to track mind wandering (Smallwood & Schooler, 2006). Before reading, participants were given the following description of mind wandering largely taken from previous studies (Smallwood & Schooler, 2006), “At some point during reading, you may realize that you have no idea what you just read. Not only were you not thinking about the text, you were thinking about something else altogether. This is called ‘zoning out.’” Participants were then informed that an auditory probe (i.e., a beep) would periodically sound during reading.
When the probe was triggered, they were instructed to hit a key marked “yes” if they were zoning out or a key marked “no” if they were not.

The auditory probes were inserted on nine pseudorandom pages for each text. The probe was triggered based on a timer, which was set to activate the probe at a random interval between 4 and 12 seconds after the participant began viewing the page. The probe was also activated if the participant attempted to advance to the next page before the timer was activated.

2.4.5. Reading times. Reading time was computed by recording the length of time participants spent reading a text minus the time it took to provide any self-reports during reading.

2.4.6. Learning measures. A knowledge test consisting of multiple-choice deep reasoning questions was used to measure learning. A deep reasoning question elicits patterns of reasoning and requires making some inference to answer the question correctly (Graesser, Ozuru, & Sullins, 2010; Graesser & Person, 1994). Each question was designed to assess knowledge of a particular concept in one of the four texts (i.e., replication, experimenter bias, causality, and dependent variables). There were 12 questions for each topic (48 questions total). Each question had four answer options. The four answer options consisted of a target (the correct response to the question), a near-miss (an option that sounds correct but was not), a thematic miss (an option that follows the theme of the content but is not actually related to the question), and a distractor (an option that is not at all related). Two versions of the test were created (24 items each) and counterbalanced as pre- and post- tests.

A transfer test was also used to assess learners’ ability to transfer acquired knowledge about scientific research methods to novel research situations. The first part of the transfer test required participants to evaluate the scientific merit of three case studies by answering four multiple-choice questions about each of the three studies, with each question being linked to a
particular topic (12 questions total). Correct responses required participants to make an inference about whether the study was flawed and what steps the researcher took to avoid potential flaws. The second portion of the transfer test presented participants with a scientific claim and asked them to design a study to test the validity of the claim. Participants were presented with eight multiple-choice options (two for each topic) that addressed different design decisions a researcher would make to design a valid study to test the claim (see Appendix A for sample items from both learning measures). The transfer test questions were directly related to the concepts covered in the four research methods text.

2.5 Procedure

Participants completed the experiment individually during a single two-hour session (see Figure 2). First, participants signed an informed consent form. They were then seated in front of a computer to begin the experiment. Next, they filled out a brief demographics survey and then completed a 24-item multiple choice pre-test. Participants were then instructed on how to use the Affect Grid using standard instructions (Russell et al., 1989) and received instructions about the auditory mind wandering probes. In order to provide a brief context for the topic of research methods, participants were given a short (< 500 words) introductory text on the broader topic of research methods followed by the baseline Topic Interest Scale (Linnenbrink-Garcia et al., 2010). Participants were then given instructions for the learning task, including the consequence value manipulation. A short, corrective multiple-choice questionnaire was presented to ensure participants understood the manipulation. That is, if they marked an answer incorrectly, the correct answer was provided immediately to clear up any misconceptions.

---

2 All learning measures were extensively piloted prior to data collection to ensure chance performance prior to reading and significant learning (but not ceiling effects) after reading. Data is available upon request.
Participants began the learning session during which the four texts were presented one at a time. The texts were displayed on a computer screen using 36 point Courier New font with 60 words per page. Participants were instructed to read each page and press the spacebar to advance to the next page. Returning to the previous page was not permitted. The consequence value of each text was presented before reading. Specifically, participants were informed that either: (a) “This text is one of the heavily weighted texts on the test. Remember, questions on this text will count three times towards your test score as compared to the regular texts” or (b) “This text has a regular weight on the test and only counts towards your test score once.”

Participants also completed an Affect Grid at four points for each text: immediately before the first page of the text, after reading one-third of the text, after reading two-thirds of the text, and upon completion of the text. A total of 9 mind wandering probes were triggered throughout each of the texts. Participants completed the Manipulation Check questions (difficulty, understanding, value, and pressure) immediately after reading each text.

Participants completed the 24-item multiple-choice posttest and the transfer test immediately after reading all four texts. Participants who did not score high enough (at least 25% or above or chance level) based on the consequence weighting (e.g., 3:1 weighting for high consequence texts) read a short text about the application of research methods (less than 500 words). The additional individual differences measures piloted during data collection were administered upon completion of the study, and did not interfere with data collection procedures or fatigue effects. Participants were fully debriefed and thanked for their participation at the end of the study.

2.6. Data Treatment
Data was collected from 187 participants. Each of the participants read four texts in the 2 × 2 (consequence value × text difficulty) design, which yielded a total of 748 (187 × 4) cases. Topic Interest was scored as participants’ average rating of interest across the ten items on the Topic Interest measure (Cronbach’s α = .828). Scores for valence (α = .977) and arousal (α = .984) were averaged across the four time points for each text. The averages of the four time points (one value per text) were used to compute Cronbach’s α for valence and arousal. Mind wandering was calculated as the proportion of “yes” responses to the nine auditory probes in each text (α = .936). Scores for the pre- and posttest were derived by computing the proportion of correct responses on the questions pertaining to each text. The average Cronbach’s alphas across the two versions of the test were .697 for the pre-test and .807 for post-test. The 20-item transfer test (α for all 20 items = .641) was scored as a weighted proportion of correct responses to the three application questions (3/5) and two design questions (2/5) relating to each of the four texts.

All Cronbach’s alphas were extremely high for the self-report measures. The Cronbach’s alphas for the learning measures were lower than the alphas for the self-reports measures, but above the “acceptable” recommendation of .6 for experimenter-generated learning measures (Loewenthal, 2001; Nunnally & Bernstein, 1994). However, it is important to note that this range of alphas was expected for the learning measures because they were constructed to assess a variety of concepts within each topic (e.g. correlation vs. causation and random assignment are distinct concepts in the Dependent Variable topic). Similar to a classroom exam, a single score is computed for each student to indicate overall performance across a wide variety of concepts and skills. Mastery of distinct concepts and skills, however, may not all be consistent since they are measuring different knowledge units.
All dependent variables were z-score standardized at the text level to remove any variance introduced by differences in content. Since data was collected at two schools with different student demographics, the dependent variables were also standardized by school in order to remove variance related to school affiliation.  

Two participants did not complete the Topic Interest measure due to computer errors. There were six cases where participants did not spend at least one minute reading a text, which was well below the average text reading time ($M = 8.15$ minutes). These six cases were removed from the data prior to the analyses.

3. Results

Unstandardized descriptive statistics for the main effects and the text difficulty $\times$ consequence value interaction are shown in Table 4 (see Appendix B for a correlation table containing all variables). A mixed-effects modeling approach (Pinheiro & Bates, 2000) using the \textit{lme4} package in R (Bates, Maechler, & Dai, 2007) was adopted for the analyses. Participant was the random effect in all models. The fixed effects were: text difficulty (easy vs. difficult), consequence value (low vs. high), and the consequence value $\times$ text difficulty interaction. Text order (order of the presentation of the texts) was a four-level categorical fixed effect to account for order and fatigue effects. Topic interest was included as a fixed effect covariate to account for general interest in the topic. Continuous variables were z-score standardized, and categorical variables were dummy coded in that one level of each categorical variable was used as the reference (or comparison) group: easy was the reference for text difficulty and low value was the reference for consequence value. Reported \textit{B} values indicate the predicted change in the dependent variable given a one unit increase in the predictor variable.

$^3$ Analyses were repeated using School and Topic as categorical fixed effects and the pattern of results did not change.
All significance testing was done with a two-tailed $\alpha$ set to 0.05. We report conservative lower-bound $F$-values and $p$-values using the \textit{LMERConvenienceFunctions} package in R (Tremblay & Ransijn, 2013). Marginal $R^2$ values, based on methods proposed by Nakagawa & Schielzeth (2013), represent the amount of variance explained by the fixed effects in each model.

3.1. Manipulation checks

It was important to ensure the manipulations of consequence value and text difficulty were successful. Participants reported higher perceived value while reading the high consequence value texts compared to the low consequence value texts, $F(1, 541) = 157.3, p < .001$, unstandardized $B = .567$, marginal $R^2 = .118$. Participants also reported higher perceived pressure while reading the high consequence value texts compared to the low consequence value texts, $F(1, 541) = 49.9, p < .001, B = .250$, marginal $R^2 = .038$. There were no significant effects for text difficulty or the consequence value $\times$ text difficulty interaction for perceived value or perceived pressure. Similarly, the models for perceptions of difficulty indicated that participants reported significantly more difficulty after reading the difficult texts compared to the easy texts, $F(1, 541) = 45.0, p < .001, B = .473$, marginal $R^2 = .073$. A similar effect of text difficulty was found for perceptions of understanding, $F(1, 541) = 23.7, p < .001, B = -.248$, marginal $R^2 = .087$. There were no differences of difficulty or understanding based on consequence value or the interaction of consequence value $\times$ text difficulty. Based on the significant main effects of consequence value and text difficulty reported in this section, we have some confidence that the experimental manipulations influenced students’ perceptions in expected directions. The remaining analyses concern the effects of the experimental manipulations on affect, attention, and learning.

3.2. The effects of consequence value and text difficulty
A direct effect of consequence value was predicted for valence and arousal, such that valence would be lower and arousal would be higher while reading the high vs. low consequence value texts. In line with these predictions, consequence value was a significant negative predictor of valence, \( F(1, 541) = 5.69, p = .017, B = -.034, \) marginal \( R^2 = .096, \) and a significant positive predictor of arousal, \( F(1, 541) = 5.83, p = .016, B = .054, \) marginal \( R^2 = .178. \) There were no direct effects of text difficulty on valence or arousal. It was also predicted that mind wandering would be less frequent for the high consequence value texts, but this was not confirmed. Instead, there was a main effect of text difficulty in that mind wandering was more frequent for the difficult texts, \( F(1, 541) = 4.18, p = .041, B = .040, \) marginal \( R^2 = .016. \) There was a marginally significant effect of consequence value on reading time with longer reading times for the high versus low value texts, \( F(1, 541) = 3.48, p = .06, B = .066, \) marginal \( R^2 = .116. \) Finally, participants also spent significantly more time reading the difficult compared to easy texts, \( F(1, 541) = 136.2, p < .001, B = .398, \) marginal \( R^2 = .116. \) The consequence value \( \times \) text difficulty interaction was not significant for valence, arousal, mind wandering, or reading time.

In addition to controlling for text order and topic interest, the models for posttest and transfer also included prior knowledge (pretest) as a covariate. Consequence value was predicted to have a direct positive effect on the posttest and transfer tests. No direct effect of consequence value was found for posttest performance. However, a significant consequence value effect was found for performance on the transfer test, \( F(1, 540) = 3.71, p = .05, B = .071, \) marginal \( R^2 = .041. \) Participants scored higher on the transfer test after reading the high consequence value texts compared to the low consequence value texts. Neither text difficulty nor the consequence value \( \times \) text difficulty effect was significant for the posttest or transfer test.

3.3 Links between affect, attention, and learning?
Two models were constructed to study if affect and attention predicted posttest and transfer performance. Valence, arousal, mind wandering, and reading time were included as fixed effects in both models. Both models also controlled for prior knowledge, topic interest, text difficulty, and consequence value. We included consequence value and text difficulty in order to determine the unique variance accounted for by valence, arousal, mind wandering, and reading time after controlling for the experimental manipulations. There were no significant relationships between posttest performance and valence, arousal, mind wandering, or reading time. However, arousal, $F(1, 537) = 6.92, p = .009, B = .058$, marginal $R^2 = .062$, valence, $F(1, 537) = 4.16, p = .042, B = .029$, marginal $R^2 = .062$, and mind wandering, $F(1, 537) = 4.63, p = .032, B = -.069$, marginal $R^2 = .062$, were all significant predictors of transfer performance. Reading time was not a significant predictor of transfer performance, $F(1, 537) = 1.20, p = .275, B = .045$, marginal $R^2 = .062$.

3.4 Mediation and moderated mediation of consequence value?

The analyses have already confirmed some of the hypothesized paths in Figure 1: (1) consequence value had a direct effect on valence, arousal, and transfer performance, and (2) valence, arousal, mind wandering predict transfer performance. Based on these links, we also tested the possibility that the effect of consequence value on transfer performance link is mediated through affect (consequence value $\rightarrow$ valence/arousal $\rightarrow$ transfer performance). We tested two mediator models: one for valence and one for arousal. Mind wandering was not tested as a mediator since there was no evidence that consequence value and mind wandering were related. Similarly, reading time was not tested as a mediator since there was no evidence that reading time was related to transfer performance.
The analysis was conducted using the ‘mediation’ package in R which has the ability to estimate multi-level mediation models (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). Average causal estimates were estimated using a quasi-Bayesian Monte Carlo method with 10,000 simulations. This estimation approach, detailed in Imai, Keele, and Tingley (2010), provides point estimates, 95% confidence intervals, and $p$-values for the mediated, direct, and total effects. The mediated effect is the average effect of consequence value on transfer performance that is due to the mediator alone. The direct effect can be interpreted as the average effect of consequence value not due to mediator and the total effect is the average combination of the mediated and direct effects. Two models are specified to assess the average causal effects: a mediator model (predicting the mediator) and an outcome variable model (predicting transfer performance). The same covariates used in all other analyses were included as fixed effects in both models as well (e.g. topic interest, prior knowledge, and order). Using the quasi- Bayesian Monte Carlo simulation method, the analyses revealed that the influence of consequence value on transfer performance was mediated by arousal (see Table 5 for a summary of the estimates, confidence intervals, and $p$-values). Thus, when consequence value was high, arousal was increased, and transfer performance was better. There was no evidence of mediation through valence.

Finally, in line with the difficulty moderation hypothesis, there is also the possibility that text difficulty may moderate the mediation effect of arousal. This type of effect, referred to as a moderated mediation, would occur if the influence of consequence value on transfer performance through arousal varied based on the level of text difficulty (Bauer et al., 2006; Preacher et al., 2007). Text difficulty and the text difficulty × arousal interaction term were added as fixed effects to both the mediator and outcome variable models. Additionally, the
outcome variable model included the text difficulty × consequence value interaction term. The mediation through arousal was assessed at the two levels of text difficulty (easy versus difficult). Indeed, results supported the difficulty moderation hypothesis since the influence of consequence value on transfer performance through arousal was only observed for the difficult texts (see Table 5 for a summary of the effects).

3.5. Summary of key findings

In sum, consequence value led to more negative valence, increased arousal, and facilitated knowledge transfer, thereby partially confirming the consequence value hypothesis. Valence, arousal and mind wandering were also significant predictors of knowledge transfer. We also found evidence of moderated mediation, such that arousal mediated the effect of consequence value on transfer performance, but only for the difficult texts. However, aside from the moderated mediation, text difficulty did not moderate any direct effects of consequence value on any of the dependent variables. Thus, the difficulty moderation hypothesis was also only partially supported.

4. Discussion

The goal of this research was to assess the influence of consequence value on affect, attention, and learning while participants learned from instructional texts on research methods and to assess if text difficulty moderated the effects. Our results suggest that increasing the consequence value of learning material indeed influenced affect and transfer performance. In addition, consequence value was found to influence transfer performance through arousal, but only for the difficult texts. In the remainder of this section, we address the extent to which the findings align with our hypotheses and discuss limitations and possible avenues of future work.

4.1. Alignment of findings with hypotheses
We first consider the degree to which our findings aligned with the multicomponential consequence value hypothesis (Hypothesis 1). Figure 3 summarizes our findings based on the links proposed by the consequence value hypothesis (see Figure 1 for hypothesized links). This hypothesis was grounded in the control-value theory (CVT), which would predict that the consequence value manipulation would be subjectively appraised as negative based on the aversive consequences associated with failure (Pekrun & Stephens, 2010). In this experiment, the manipulation was based on staying longer to read additional text, which is a negative outcome for the average undergraduate student. According to the first prediction of the consequence value hypothesis, valence was expected to be lower for the high versus low consequence value texts due to an increased focus on the potential negative outcome. Self-determination theory (SDT) would also predict lower valence in this situation since consequences reduce autonomy and lower intrinsic motivation. This prediction was confirmed in the present experiment as participants reported lower valence while reading the high versus low consequence value texts.

Although SDT and CVT make a similar prediction for valence, CVT addresses how consequence value influences arousal. Specifically, CVT predicts that valence is not the only affective dimension of interest, but that arousal is an equally important dimension to consider. We also predicted arousal would be higher for the high consequence value texts due to a focus on the threat of failure and its negative consequences. The results also supported this part of the hypothesis. Not only did participants report lower valence for the high consequence value texts, they also reported higher arousal. We consider the lower valence and higher arousal for the high vs. low consequence value texts to reflect a state of mild anxiety induced by the consequence value manipulation. Participants took longer to read the high consequence value texts, supporting the prediction that students exhibited more effort while reading the high value texts.
The prediction that mind wandering would occur less during the high versus low value texts was not supported. The portion of the hypothesis was based on the idea that feelings of mild anxiety during the high consequence value texts would be related to increased attentional focus. Instead, our findings suggest that consequence value did not influence mind wandering. There was also no evidence to suggest that the consequence value manipulation led to overall lower rates of mind wandering, as the rates of mind wandering found in this study (29%) were on par with previous research with no consequence value manipulation (Feng, D’Mello, & Graesser, 2013; Smallwood et al., 2007; Smallwood & Schooler, 2006).

The consequence value hypothesis also predicted that high consequence value would positively influence learning, measured here via performance on the posttest and transfer test (see Figure 1). However, this prediction was only partially supported, as consequence value led to higher performance on the transfer test, yet had no effect on posttest performance. The same pattern was found for the influence of valence, arousal, and mind wandering on learning, such that transfer performance (and not posttest) was positively related to valence and arousal, and negatively related to mind wandering. Based on the links shown in Figure 1, arousal and mind wandering predicted transfer performance in the expected directions. No direction was originally predicted for the relationship between valence and learning based on previously mixed findings (Fielder & Beier, 2014; Isen et al., 1987). However, a positive relationship was found in the current study.

Based on the confirmed links between consequence value, affect, attention, and learning (see Figure 1), there was also a question of whether affect or attention mediated the relationship between consequence value and transfer performance. The positive effect of consequence value on transfer performance was mediated by arousal, but not through valence or mind wandering.
Importantly, higher consequence value led to elevated levels of arousal, presumably due to a focus on the potential for failure. It was through the increased arousal that participants performed better on the transfer test.

In sum, the consequence value hypothesis was partially supported in that: (1) consequence value influenced valence and arousal, and had a marginal effect on reading time but not mind wandering, (2) affect (valence and arousal) and attention (mind wandering) predicted transfer performance, and (3) the effect of consequence value on transfer performance was mediated by arousal.

Next we consider the extent to which text difficulty moderated the effects of consequence value as predicted by the difficulty moderation hypothesis (Hypothesis 2). We found no evidence to suggest that text difficulty moderated any direct effects of consequence value on affect, attention, or learning. One potential reason for this pattern of results is based on the overall strength of the consequence value manipulation. Participants reported increased arousal, more negative valence, and had higher transfer performance during the high value texts, regardless of whether the text was easy or (moderately) difficult to read. This perhaps indicates the moderate level of text difficulty was an inconsequential factor with respect to directly influencing affect or transfer performance.

Aside from considering how text difficulty may have influenced the direct effects of consequence value, we also considered whether text difficulty moderated the mechanism by which consequence value influenced transfer performance through arousal. The difficulty moderation hypothesis predicted that the influence of consequence value through affect and attention may only occur when texts are moderately difficult compared to when they are easy. In particular, difficult texts may require increased arousal as a mechanism for learning to a different
extent in comparison to the easy texts. It is important to point out that this portion of the hypothesis could be tested despite the fact that we found no evidence that text difficulty moderated the direct effects of consequence value. Moderated mediation need not be dependent on an interaction between the independent variable and moderator, even though the interaction term is nevertheless included in the model (Fairchild & MacKinnon, 2009).

A moderated mediation analysis indicated that the mediating effect of arousal only occurred when the texts were difficult and not for the easy texts. This suggests that the path from consequence value to learning (consequence value $\rightarrow$ arousal $\rightarrow$ transfer performance) only necessitated arousal when the text was difficult. Thus, when confronted with a text that required more effortful processing (e.g., difficult), increased levels of arousal served as a mechanism for improved transfer performance. In sum, text difficulty did not moderate the direct effects of consequence value, yet was a moderator of the mediation through arousal. The moderated mediation analysis provides evidence that arousal may play a facilitative role in some, but not all, circumstances.

Although we did not predict any direct effects of text difficulty, text difficulty had a positive effect on reading time and mind wandering. Difficult texts were specifically manipulated to target factors known to take longer to process (e.g., infrequent words and complex syntax; Graesser, Hoffman, & Clark, 1980; Graesser et al., 2011), which may explain participants’ increased reading times while reading these texts. Mind wandering was also reported more frequently during the difficult texts. This effect replicates previous research which found more mind wandering when reading difficult versus easy texts (Feng et al., 2013; Mills et al., 2013). Despite the increase in mind wandering, however, text difficulty did not have a negative influence on learning on either the posttest and transfer test. However, mind
wandering was found to negatively influence transfer performance. Based on these findings, it is possible that mind wandering may have mediated the relationship between text difficulty and transfer performance. We therefore performed a post-hoc analyses to test this mediation model (text difficulty $\rightarrow$ mind wandering $\rightarrow$ transfer performance), but there was no evidence of mediation through mind wandering.

Finally, it is also important to acknowledge that neither of the experimental manipulations influenced posttest performance, nor was posttest performance related to affect or attention. One potential reason for the differential patterns between the posttest and transfer test is that both tests tapped different types of knowledge. The transfer test required higher-level inferences about the research methods topics in comparison to the posttest, making the transfer test more difficult. This is evidenced by the significantly lower scores on the transfer test ($M = .497, SD = .187$) compared to the posttest ($M = .611, SD = .199$), paired $t(187) = 11.15, p < .001$. In addition, the two performance tests were only moderately correlated ($r = .356$) after controlling for pre-test performance. Therefore, the high consequence value manipulation may have influenced learning through a deeper understanding of research methods concepts, but only for the transfer test, as the posttest scores were higher regardless of consequence value.

4.2. Limitations and future directions

It is important to note a few limitations with the current experiment. First, this was a lab study with a sample of 187 college students, so it is important to interpret the findings with a modicum of caution as affect and attention be influenced differently outside of the laboratory. Second, although we attempted to control for topic interest and prior knowledge, we did not account for other trait-level variables that may have influenced affect and attention (e.g., language skills and working memory capacity). Third, the text difficulty manipulation was
limited to two experimenter-generated levels of difficulty. Since text difficulty did not influence learning, the difficult texts may not have been particularly challenging for the participants in this study. Thus, one suggestion for future studies is to manipulate difficulty is by adjusting difficulty levels based on each individual’s abilities (Fulmer & Tulis, 2013). Another alternative would be to manipulate levels of subjective control using a different method altogether, perhaps by increasing the time constraints of the learning activity. Fourth, a (no consequence value) control condition was not included in this experiment. A no-value control condition would have allowed us to determine the effects of imposing any value incentive at all, even compared to a low consequence value condition. Finally, reading was periodically interrupted for participants to provide reports (via affect grid or responses to mind wandering probes) and these interruptions could have interfered with the primary learning activities. Thus, less intrusive measures, such as retrospective affect assessments and mind wandering reports may be warranted in future research.

4.3. Concluding remarks

Previous research has shown that consequences can have both positive and negative effects on motivation and performance. The present experiment attempted to reconcile the conflicting findings by investigating the influence of consequence value on two different, yet vitally important processes – affect and attention during learning. The present results suggest that while consequence value can lead to negative affect, it may also increase arousal which can facilitate deeper learning. In particular, arousal mediated the effect of consequence value on transfer performance for the difficult texts. This is significant since motivating students can be a daunting task, especially for topics that are not particularly interesting to them. This study provides an example of one context where extrinsically-valued consequences may be beneficial
in promoting productive affective states during an instructional learning activity. As a next step, these ideas should be tested in a classroom environment to see if similar patterns emerge in a more ecological context. A better understanding of the factors that can influence students’ affect and attention will be important in the development of empirically-based motivational strategies that can be used in the classroom.
Acknowledgements

This research was supported by the National Science Foundation (NSF) (DRL 1235958). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF. We would like to thank Blair Lehman, Shi Feng, and Andrew Olney at the University of Memphis for their help with data collection.
References


http://doi.org/10.2307/747086


http://doi.org/10.1177/0013164409355699


Table 1

*Example 2 × 2 Overview of Subject Value and Levels of Arousal based on CVT*

<table>
<thead>
<tr>
<th>Arousal</th>
<th>Negative Subjective Value</th>
<th>Positive Subjective Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate Arousal</td>
<td><strong>Mild Anxiety</strong></td>
<td>Engagement</td>
</tr>
<tr>
<td>Low Arousal</td>
<td>Boredom/Disengagement</td>
<td>Relaxation/Calmness</td>
</tr>
</tbody>
</table>
Table 2

*Examples showing difficulty manipulations*

<table>
<thead>
<tr>
<th>Manipulation</th>
<th>Easy</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Length</td>
<td>It is possible that in reality people really have no taste preference for free-range chickens. The people in Dr. Jones' study just happened to give the free-range chicken meat higher ratings by chance. In other words, it could just be that the results in the original study were wrong. <em>(3 short sentences)</em></td>
<td>It is possible that people in actuality have no taste preference for free range chickens but the ones in her study just happened to give the free range chicken meat higher ratings by chance and it could just be that the results in the original study were wrong. <em>(1 long sentence)</em></td>
</tr>
<tr>
<td>Word Frequency</td>
<td>This means that <em>skipping</em> sleep one night will cause them to get <em>bad</em> grades.</td>
<td><em>Relinquishing</em> sleep one night will have <em>deleterious</em> effects on their grades.</td>
</tr>
<tr>
<td>Verb Cohesion</td>
<td>Scientists do not <em>talk about</em> the truth of a theory. This is because the truth can never really be known. Instead, they <em>talk about</em> how likely a theory is to be true.</td>
<td>Scientists don't <em>strive to aver the truth</em> of a theory because it can never in reality be known; instead they <em>talk about</em> how likely it is that the theory is spot-on.</td>
</tr>
<tr>
<td>Syntactic complexity</td>
<td>We would need to find out if more people stop smoking after they <em>listened to the subliminal tape</em> than another group of people who <em>did not listen to the tape</em>.</td>
<td>We would need to discover whether more individuals stop smoking after <em>listening to the tape more so</em> than another group who <em>also wanted to stop smoking</em> but did not listen to it.</td>
</tr>
</tbody>
</table>
Table 3

*Means and standard deviations (in parenthesis) of textual features between easy and difficult texts*

<table>
<thead>
<tr>
<th>Text Level Manipulations</th>
<th>Easy Texts $M$ $(SD)$</th>
<th>Difficult Texts $M$ $(SD)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb Cohesion</td>
<td>41.3 (7.96)</td>
<td>19.4 (6.64)</td>
</tr>
<tr>
<td>Syntactic Simplicity</td>
<td>82.6 (4.74)</td>
<td>42.7 (4.07)</td>
</tr>
<tr>
<td>Number of Words per Sentence</td>
<td>14.2 (.551)</td>
<td>25.6 (1.26)</td>
</tr>
</tbody>
</table>

| Estimated Grade Levels                        |                       |                             |
| Flesch-Kincaid Grade Level                    | 7.09 (.237)           | 13.06 (.417)                |
| Pearson Reading Maturity                      | 8.20 (.638)           | 13.28 (.556)                |
| Automated Readability Index                   | 7.20 (.668)           | 14.73 (.978)                |
Table 4

*Unstandardized means and standard deviations (in parenthesis) for main effects and interactions*

<table>
<thead>
<tr>
<th>Measures</th>
<th>Low</th>
<th>High</th>
<th>Easy</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consequence Value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value Perception [1-6]</td>
<td>3.63 (1.26)</td>
<td>4.51 (1.13)</td>
<td>4.12 (1.17)</td>
<td>4.05 (1.18)</td>
</tr>
<tr>
<td>Pressure Perception [1-6]</td>
<td>2.56 (1.39)</td>
<td>2.95 (1.51)</td>
<td>2.70 (1.45)</td>
<td>2.81 (1.47)</td>
</tr>
<tr>
<td>Difficulty Perception [1-6]</td>
<td>2.38 (.923)</td>
<td>2.48 (.930)</td>
<td>2.19 (.890)</td>
<td>2.67 (1.00)</td>
</tr>
<tr>
<td>Understand Perception [1-6]</td>
<td>4.12 (1.21)</td>
<td>4.04 (.123)</td>
<td>4.22 (1.16)</td>
<td>3.94 (1.26)</td>
</tr>
<tr>
<td><strong>Affect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence [1-9]</td>
<td>4.83 (1.36)</td>
<td>4.71 (1.38)</td>
<td>4.83 (1.32)</td>
<td>4.72 (1.42)</td>
</tr>
<tr>
<td>Arousal [1-9]</td>
<td>3.87 (1.60)</td>
<td>4.04 (1.66)</td>
<td>4.00 (1.63)</td>
<td>3.91 (1.59)</td>
</tr>
<tr>
<td><strong>Attention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. MW [0-1]</td>
<td>.284 (.326)</td>
<td>.297 (.331)</td>
<td>.278 (.326)</td>
<td>.305 (.331)</td>
</tr>
<tr>
<td>Reading Time [minutes]</td>
<td>8.15 (2.53)</td>
<td>8.28 (2.51)</td>
<td>7.73 (2.16)</td>
<td>8.69 (2.76)</td>
</tr>
<tr>
<td><strong>Knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. Pre [0-1]</td>
<td>.523 (.202)</td>
<td>.506 (.194)</td>
<td>.511 (.194)</td>
<td>.520 (.203)</td>
</tr>
<tr>
<td>Prop. Post [0-1]</td>
<td>.607 (.210)</td>
<td>.622 (.217)</td>
<td>.619 (.208)</td>
<td>.614 (.221)</td>
</tr>
<tr>
<td>Prop. Transfer [0-1]</td>
<td>.433 (.234)</td>
<td>.469 (.236)</td>
<td>.466 (.245)</td>
<td>.436 (.250)</td>
</tr>
</tbody>
</table>

*Notes.* Values of significant differences appear in bold font ($p < .05$). Range of measures in brackets; Understand = Understanding; Prop. MW = Proportion of probes reported “yes” to mind wandering; Prop. Pre = Proportion correct on pre-test; Prop. Post = Proportion correct on post-test; Prop. Transfer = Proportion correct on transfer test.
Table 5

*Mediation by arousal estimate, 95% confidence intervals, and p-value for overall mediation and moderated mediation (split by easy and difficult)*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Overall Mediation by Arousal</th>
<th>Moderated Mediation by Arousal</th>
<th>Easy Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est. 95% CI  p</td>
<td>Difficult Texts Est. 95% CI  p</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediation</td>
<td>.01  .001; .0151 .04</td>
<td>.01  .0016; .0315 .02</td>
<td>.00 -.0026; .0154 .26</td>
</tr>
<tr>
<td>Direct</td>
<td>.09  -.0048; .1918 .06</td>
<td>.08  -.0133; .1834 .09</td>
<td>.08 -.0133; .1834 .09</td>
</tr>
<tr>
<td>Total</td>
<td>.10  .0013; .1982 .05</td>
<td>.09  -.0011; .1978 .05</td>
<td>.08 -.0102; .1878 .08</td>
</tr>
</tbody>
</table>

*Notes. 95% CI = 95% confidence intervals; Est. = estimate; Bold values = 95% CI does not contain zero.*
Figure 1. Diagram of hypothesized relationships.
Figure 2. Experimental protocol flowchart.
Figure 3. Diagram of supported relationships based on hypothesized links in Figure 1. * $p < .10$. 
Appendix A

Sample Items from Learning Assessments

Sample item from posttest

Which scenario would be the best to use a double blind study?

a) Researchers are testing the relationship between test anxiety and GPA (thematic miss)

b) **Researchers are testing the effects of lotion on making people look younger (correct answer)**

c) Researchers are testing the effect of a new soap on bacteria cells (near miss)

d) None of the above (distractor)

Sample item from transfer test

A sports psychologist tested whether visualizations alone have an impact on muscle tone. Visualization involves picturing an activity (e.g., shooting a basketball) in your mind, but not physically doing the activity. Earlier studies compared physical exercise only to a group that did physical and visualization exercise, but the sports psychologist added two groups to her study: (1) visualization exercise only and (2) a control group that did neither. All participants were first given a standardized test of muscle tone (biceps) by a trained professional. Participants were randomly assigned to one of the four conditions. For four weeks, participants (except the control group) spent 20 minutes every morning performing exercises (physical, visualized, or physical and visualized) designed to strengthen biceps. After four weeks, bicep muscle tone was measured by the same expert, unaware of participants’ condition. Both physical exercise and visualization exercise had significant effects on muscle tone, confirming earlier results that visualization exercise effectively increases muscle tone. The researcher repeated the study and found the same results.

The conclusions being drawn by the researchers are...

a. Causal because the researcher measured muscle tone before and after the treatment conditions.

b. Correlational because the researcher measured muscle tone before and after the treatment conditions.

c. **Causal because the participants were randomly assigned to conditions (correct answer)**

d. Correlational because the participants were randomly assigned to conditions.
Appendix B

Subject–level correlations between standardized variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Value Percp.</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Pressure Percp.</td>
<td>.317</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Difficulty Percp.</td>
<td>.084</td>
<td>.358</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Understand Percp.</td>
<td>.456</td>
<td>.092</td>
<td>-.294</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Reading Time</td>
<td>.295</td>
<td>.174</td>
<td>.040</td>
<td>.175</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Valence</td>
<td>.127</td>
<td>-.102</td>
<td>-.056</td>
<td>.219</td>
<td>.013</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Arousal</td>
<td>.185</td>
<td>.193</td>
<td>.119</td>
<td>.264</td>
<td>.080</td>
<td>.274</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Prop. MW</td>
<td>-.286</td>
<td>-.026</td>
<td>.103</td>
<td>-.338</td>
<td>-.017</td>
<td>-.184</td>
<td>-.189</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) Prop. Pre</td>
<td>.164</td>
<td>-.025</td>
<td>-.104</td>
<td>.261</td>
<td>-.117</td>
<td>.150</td>
<td>.166</td>
<td>-.169</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) Prop. Post</td>
<td>.286</td>
<td>.034</td>
<td>-.113</td>
<td>.398</td>
<td>.052</td>
<td>.163</td>
<td>.158</td>
<td>-.182</td>
<td>.525</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(11) Prop. Transfer</td>
<td>.254</td>
<td>.012</td>
<td>-.013</td>
<td>.366</td>
<td>-.014</td>
<td>.171</td>
<td>.250</td>
<td>-.268</td>
<td>.479</td>
<td>.513</td>
<td>1</td>
</tr>
<tr>
<td>(12) Topic Interest</td>
<td>.173</td>
<td>.057</td>
<td>.064</td>
<td>.144</td>
<td>-.017</td>
<td>.292</td>
<td>.394</td>
<td>-.100</td>
<td>.137</td>
<td>.270</td>
<td>.168</td>
</tr>
</tbody>
</table>

Notes. Values of significant differences appear in bold font (p < .05). Percp. = Perception; Understand = Understanding; Prop. MW = Proportion of probes reported “yes” to mind wandering; Prop. Pre = Proportion correct on pre-test; Prop. Post = Proportion correct on post-test; Prop. Transfer = Proportion correct on transfer test.