Contemporary Educational Psychology 41 (2015) 98–110

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Interest-based text preference moderates the effect of text difficulty on engagement and learning

Sara M. Fulmer [a,*], Sidney K. D'Mello [b], Amber Strain [c], Art C. Graesser [c]

[a] Department of Educational Psychology, Counseling, and Special Education, SUNY Oneonta, Oneonta, NY, 13820, USA
[b] Departments of Psychology and Computer Science, University of Notre Dame, Notre Dame, IN 46556, USA
[c] Department of Psychology, University of Memphis, Memphis, TN 38152, USA

ARTICLE INFO

Article history:
Available online 24 December 2014

Keywords:
Interest
Difficulty
Engagement
Learning

ABSTRACT

Theories of motivation propose that moderate difficulty can be beneficial for student engagement and learning. However, research on the effect of difficulty has been inconsistent. The primary goal of the present study was to investigate the possibility that interest-based text preference moderates the effect of difficulty on engagement and learning. To test this hypothesis, participants studied four instructional texts on research methods topics in a 2 × 2 interest-based text preference (preferred vs. non-preferred texts) × text difficulty (easy vs. difficult) within-subjects experiment. The manipulation of interest-based text preference asked participants to rank four text titles based on their perceived interest in reading the text corresponding to each title. Engagement was assessed via self-reported affect (valence and arousal), attention (mind wandering), and reading time during the learning session. Learning and knowledge transfer were measured with knowledge tests after reading all four texts. Consistent with our predictions, interest-based text preference and text difficulty interacted to predict reading time, mind wandering, and knowledge transfer. The nature of the relationship indicated that increased text difficulty can support engagement and transfer, but only when individuals are provided with an opportunity to express their text preferences prior to reading.

1. Introduction

Most individuals have experienced either momentary lapses in attention or sustained periods of disengagement during learning. These declines in engagement not only have the immediate consequence of impeding comprehension, but if sufficiently frequent, can have undesirable long-term consequences such as a lack of interest in the particular domain. What characteristics of the learning activity inspire sustained engagement and deep learning, versus zoning out, quitting, and negligible learning? Theories of motivation posit that moderate difficulty has direct, positive effects on engagement and learning (Csikszentmihalyi, 1991; Deci & Ryan, 1985). However, as discussed below, empirical research on the independent effects of difficulty on engagement has been inconsistent. Thus, research is needed to identify the specific conditions under which increased difficulty helps, hurts, or has no effect.

There is a long history of research on the importance of autonomy and interest for increasing intrinsic motivation, thereby supporting increased engagement and learning (Ainley, 2012; Ainley & Hidi, 2014; Deci & Ryan, 1985; Hidi, 1990; Hidi & Renninger, 2006; Schiefele, 1991). This led us to consider the possibility that autonomy and interest, operationalized as an interest-based text preference, would serve as a motivational resource (Hidi, 1990) to support engagement and learning while reading difficult texts. This paper tested this hypothesis by investigating the interaction between manipulations of interest-based text preference and text difficulty during learning from instructional texts. The manipulation of interest-based text preference consisted of asking participants to indicate which texts they would like to read based on their perceived interest after reading the text titles. The text difficulty manipulation included experimenter-created easy and difficult versions of the texts.

In what follows, we first discuss the operationalization of engagement adopted in the present study. We then review the literature on the effects of text difficulty and interest-based preference on engagement and learning, and our rationale for hypothesizing that interest-based text preference will moderate the effects of text difficulty on engagement and learning. In the following review, we use the general term *difficulty* to refer to research that has manipulated the objective difficulty of the task. The specific term *text difficulty* refers to studies that have manipulated the difficulty level of texts, similar to the present study.

* Corresponding author. SUNY Oneonta, 365 Fitzelle Hall, Oneonta, NY 13820, USA. Fax: +607-436-3664.
E-mail address: sara.fulmer@oneonta.edu (S.M. Fulmer).

http://dx.doi.org/10.1016/j.cedpsych.2014.12.005
0361-476X/

1.1. Operationalizing engagement

Engagement is a multidimensional construct that has been used to describe diverse behaviors, thoughts, perceptions, feelings, and attitudes (Reschly & Christenson, 2012). Fredricks, Blumenfeld, and Paris (2004) proposed three types of engagement: emotional, behavioral, and cognitive. Generally speaking, emotional engagement encompasses affective states that are experienced during learning, including mood, affect, achievement emotions, epistemic emotions, and interest (Fredricks et al., 2004; Pekrun & Linnenbrink-Garcia, 2012). Behavioral engagement is broadly defined as learners' participation and involvement in a learning task, including their effort, persistence, and concentration (Fredricks et al., 2004; Fredricks & McColskey, 2012). Finally, cognitive engagement is related to learners' investment in the task, such as how they manage and control effort directed toward learning, understanding, and mastery of the material (Newmann, Wehlage, & Lamborn, 1992; Pintrich & De Groot, 1990; Zimmerman, 1990). Recent conceptualizations of engagement further distinguish cognitive engagement (e.g., attention and memory processes) from cognitive–behavioral engagement (e.g., strategy use and self-regulation) (Pekrun & Linnenbrink-Garcia, 2012), suggesting that there is some conceptual overlap between cognitive and behavioral engagement (Fredricks et al., 2004). Accordingly, there is little consensus regarding how to operationalize and measure each aspect of engagement, so multiple measures are often recommended (Shernoff, 2013).

Research on academic engagement has traditionally focused on the antecedents that support or suppress engagement in learning contexts that span extended periods of time (Christenson, Reschly, & Wylie, 2012; Linnenbrink-Garcia & Pekrun, 2011; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). However, contemporary views of engagement suggest that engagement can fluctuate during learning in ways that are sensitive to the context and features of the task. Pursuant to this view, research has moved beyond exclusive "trait" (or single time point) measures of engagement to on-task, "state" measures in order to account for its temporal, dynamic, and context-sensitive nature (Ainley, Corrigan, & Richardson, 2005; Pekrun & Schutz, 2007; Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003). State engagement is defined as a state of involvement with a task that is characterized by mild positive valence, moderate arousal, and intense, focused attention (Baker, D'Mello, Rodrigo, & Graesser, 2010). It is related to, but need not involve, some aspects of Csikszentmihalyi's (1991) conceptualization of flow, such as time distortion or loss of self-consciousness. The present study adopts such a conceptualization of engagement and studied it during a shorter task, which involved learning from instructional texts for approximately 30 minutes.

1.2. Theoretical background and previous research on difficulty and interest-based preferences

There is extensive theoretical support for moderate difficulty as a favorable condition for engagement, deep cognitive processing, and learning. For example, early theories of risk-taking and exploration (Atkinson, 1957; Berlyne, 1954) suggested that individuals are more likely to seek out tasks that are unfamiliar, complex, or novel, presumably when intrinsic interest is high. These task features increase interest and curiosity because of their potential for learning something new, leading to greater arousal, persistence, and retention of information. Self-determination theory proposes a positive relationship between moderate difficulty and intrinsic motivation, such that moderate difficulty that satisfies the needs for competence and autonomy should enhance intrinsic motivation and, consequently, engagement (Deci & Ryan, 1985; Ryan & Deci, 2000). Moderate difficulty is also proposed to increase positive affect, interest, persistence, and attention (Clifford, 1990; Csikszentmihalyi, 1991; Wigfield & Eccles, 2002).

When tasks are not intrinsically motivating, it has been suggested that introducing *desirable difficulties* into the learning context can inspire deeper cognitive processing and result in greater learning and long-term retention (Bjork & Bjork, 2011; Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Linn, Chang, Chiu, Zhang, & McElhaney, 2011). Examples of desirable difficulties include using distributed practice, varying the settings in which learning occurs, and presenting learning materials with minor disfluencies. The general idea is that people are "cognitive misers" in that they allocate just the appropriate amount of cognitive effort to a task. The desirable difficulties address this by reconfiguring the task context to engender deeper processing (Alter, 2013). Despite the strong theoretical support for a positive relationship between difficulty and engagement, research on difficulty is rife with inconsistent findings. For example, learners spend more time reading difficult texts by calibrating effort with respect to difficulty (Feng, D'Mello, & Graesser, 2013; Vega, Feng, Lehman, Graesser, & D'Mello, 2013). However, higher perceived difficulty of informational texts predicts greater aversion to reading informational texts, suggesting a decline in behavioral engagement with increasing perceived difficulty (Guthrie, Klauda, & Ho, 2013). Similarly, increased mind wandering (or zone outs) has been reported in easy versions of lab-based cognitive tasks (Smallwood & Schooler, 2006), but also when reading difficult compared with easy texts (Feng et al., 2013). Though more difficult texts (in terms of cohesion) provide opportunities for deeper learning, readers must be sufficiently vested and have the relevant background knowledge and skill to utilize effective reading strategies in order for these texts to enhance comprehension (McNamara & Kintsch, 1996). Finally, with respect to learners' interest and emotions, higher perceptions of difficulty have been related to lower self-reported interest (Durik & Matarazzo, 2009; Li, Lee, & Solmon, 2007), lower levels of happiness (Moneta & Csikszentmihalyi, 1996), and an increase in self-reported affective states involving negative valence, such as anxiety, anger, and boredom (Acee et al., 2010; Efklides, 2002; Efklides & Petkaki, 2005; Pekrun, Goetz, Titz, & Perry, 2002).

What is the reason for these contradictory findings? It might be the case that the proposed benefits of difficulty for engagement and learning assume that learners calibrate their attention and effort with respect to the level of difficulty. However, learners must decide whether or not to exert this additional effort, and are unlikely to do so in the absence of intrinsic motivation. The present study considers the expression of learners' interests via their text preferences as a potential motivational resource (Hidi, 1990) to foster engagement and deeper learning while studying difficult texts.

The favorable role of interest during learning has support in self-determination theory (Ryan & Deci, 2000), which posits that feeling autonomous during learning increases intrinsic motivation, which leads to enhanced performance and learning. Individuals feel autonomous when they perceive that their actions and the outcomes of these actions originate from, and are endorsed by, the self (Deci, Vallerand, Pelletier, & Ryan, 1991). Additionally, autonomy-supportive learning environments are those that acknowledge learners' inner motivational resources, including their personal interests, values, and preferences (Reeve & Jang, 2006). In line with this, the present study aims to support feelings of autonomy by offering learners the opportunity to express their preferences for texts based on their personal interests (termed "interest-based text preference"). The opportunity to communicate one's interests and preferences supports an internal locus of control by acknowledging one's inner motivational resources during the learning activity (Deci & Ryan, 1985; Krapp, 2002; Reeve & Tseng, 2011), and is expected to lead to greater engagement and deeper learning.

In general, interest is an important motivational resource for boosting and sustaining engagement during unfavorable learning activities (e.g., reading difficult texts) through increasing feelings of arousal and attention (Ainley, 2006; Ainley & Hidi, 2014; Ainley, Hidi, & Berndorff, 2002; Alexander & Jetton, 1996; Hidi, 1990; Hidi & Renninger, 2006; Linnenbrink-Garcia, Patall, & Messersmith, 2013; Schiefele, 1991). For example, higher topic interest in a text is related to higher enjoyment and greater persistence during excessively difficult texts among middle school students (Fulmer & Frijters, 2011). These findings prompted the hypothesis for the current study that interest-based text preference may moderate the effects of text difficulty on engagement and learning while reading instructional texts.

1.3. Present study

It is important to discuss how engagement was measured in the present study before delving into the specific hypotheses and study design. The present study included three measures to assess various aspects of engagement: affective states (valence and arousal) as a measure of affective engagement, and reading time and mind wandering as measures of cognitive/behavioral engagement. Participants' affective states were assessed by their self-reported levels of valence and arousal during the learning task. In particular, the present study uses a circumplex model of affect (Linnenbrink, 2007; Russell, 1980), which posits that affective states are characterized by two orthogonal dimensions: valence and arousal (also termed activation). The valence dimension differentiates positive (e.g., excitement) from negative (e.g., sadness) affective states, while the arousal dimension differentiates activating (e.g., excitement) from deactivating (e.g., relaxation) states (Shuman & Scherer, 2014). For example, excitement is a positive, activating affective state, while relaxation is a positive, deactivating affective state. Similarly, negative affective states include activating states (e.g., anger) and deactivating states (e.g., sadness). Thus, as an alternative to assessing specific emotions, the circumplex model provides a comprehensive perspective on how learners' affective states vary along these two dimensions.

The second measure of engagement was the amount of time participants spent on a self-paced learning task. When taken as an absolute measure, it is unclear if engagement should be reflected via faster (more concentration) or slower (more elaborative processing) reading times. Furthermore, absolute reading time is likely conflated with reading ability in that better comprehenders should read faster irrespective of engagement. Thus, the present analysis focuses on *relative* rather than *absolute* reading times by comparing reading times of the same participant for the easy vs. difficult texts. We expect that participants who are engaged should spend more time reading the difficult versus easy texts because we adopted difficulty manipulations that target factors (e.g., more complex syntax and lower frequency words) which are known to take longer to process if attended to (Beck, McKeown, & Kucan, 2013; Graesser & McNamara, 2011; Haberlandt & Graesser, 1985).

The third measure of engagement was participants' tendency to sustain attention during learning by suppressing mind wandering behaviors. Mind wandering (sometimes referred to as zoning out or daydreaming) is a phenomenon that occurs when attention automatically shifts away from the processing of task-relevant information to the processing of internal, task-irrelevant thoughts (Giambra, 1995; McVay & Kane, 2012; Smallwood, Fishman, & Schooler, 2007). Although some types of task-relevant mind wandering can be beneficial for content integration and creative thinking (Mooneyham & Schooler, 2013), we investigated attention reallocation away from the text to task-irrelevant thoughts (e.g., daycare, dinner, an upcoming vacation). This type of mind wandering is an indicator of lower cognitive and behavioral engagement because learners are no longer focusing on the content of the text and are unable to self-regulate and avoid internal distractions (Kane & McVay, 2012). Consequently, this form of mind wandering negatively impacts the encoding of information (Seibert & Ellis, 1991) and comprehension of texts (Feng et al., 2013; McVay & Kane, 2012; Schooler, Reichle, & Halpern, 2004).

To test the interaction between interest-based text preference and text difficulty, we manipulated two aspects of the task in a 2×2 within-subjects design: the provision of interest-based text preference (preferred vs. non-preferred texts) and text difficulty (easy vs. difficult). Given our primary interest in testing the interaction between these two constructs, we refrain from making any hypotheses on their direct effects on engagement and learning. Instead, we hypothesize that text difficulty effects will vary based on interest-based text preference, as detailed below.

For the non-preferred texts, we hypothesize that engagement and learning will be lower when reading the difficult compared with easy texts. Non-preferred texts are the texts that participants explicitly indicate as not wanting to read. These texts are expected to reduce learners' feelings of autonomy, making it less likely that they will actively engage with the texts. As a result, learners will exert minimal effort toward comprehending the texts. Low effort and shallow processing might suffice for comprehending easy texts, but will be inadequate for difficult texts, which require the learner to continually exert effort in order to sustain engagement with the texts and support learning.

For the preferred texts, we expect one of two outcomes: (1) engagement and learning will be equivalent for easy and difficult texts; or (2) engagement and learning will be higher during difficult texts. In particular, if the advantage of interest-based text preference during learning is that it *only* helps the learner to manage the potential negative influence of difficulty (Renninger, Ewen, & Lasher, 2002), then engagement and learning will be equivalent for preferred, easy texts and preferred, difficult texts. However, if there is an *added* advantage of interest-based text preference for engagement and learning (Fulmer & Frijters, 2011), such that it inspires greater attention, concentration, and deeper processing, we expect that participants will report higher engagement and demonstrate greater learning for difficult texts, compared with easy texts. Based on our operationalization of engagement, this higher engagement will be reflected in more positive-activating affect (positive valence and moderate arousal), less mind wandering, and greater reading time for the difficult compared with the easy texts.

2. Methods

2.1. Participants

Participants were 84 undergraduate students ($M_{age} = 23.17$, $SD = 6.51$, 72.6% female) enrolled in the Psychology department from a large public university in the mid-South United States. Participants' age ranged from 18 to 49 years. The majority of participants were African-American (53.6%), with 23.8% Caucasian, 7.1% Asian, 2.4% Hispanic, and 12.1% non-responders. Seventy-five participants participated for undergraduate course credit and the remaining nine participants received monetary compensation for their participation.

2.2. Experimental manipulations

This experiment used a 2×2 (text difficulty $\times$ interest-based text preference) within-subjects design. The order of conditions and topics and assignment of topic to condition was counterbalanced across participants with a Latin Square.

2.2.1. Instructional texts

Four texts on research methods concepts were obtained from the electronic textbook that accompanies the educational game

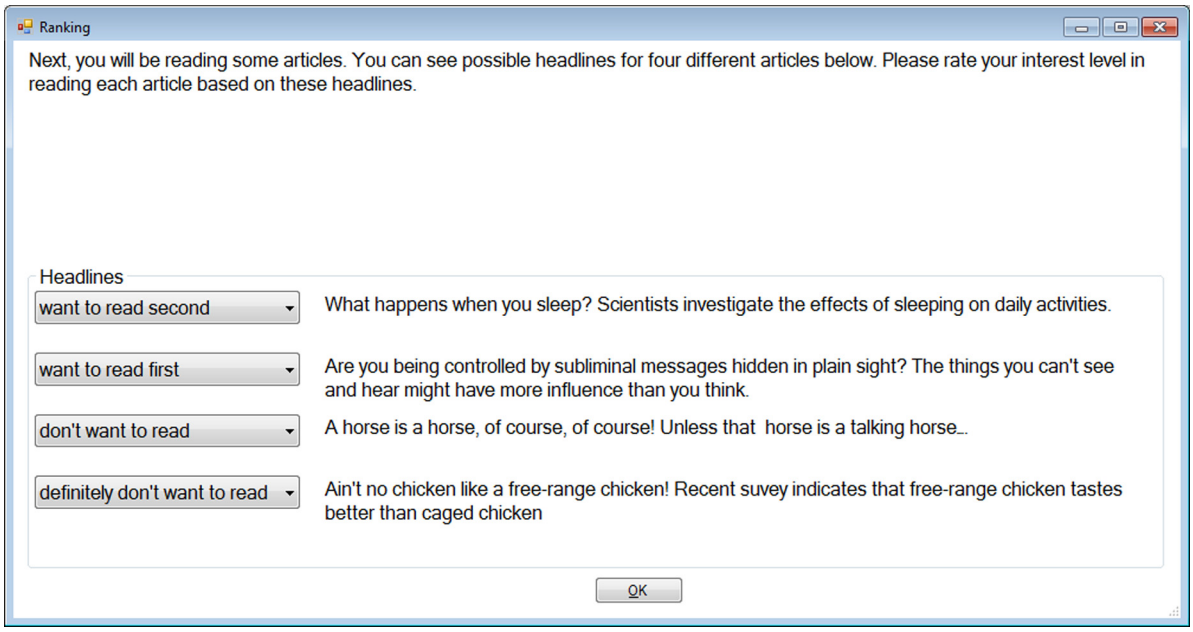

Fig. 1. Screenshot of topic ranking interface and text titles.

Operation ARA! (Halpern et al., 2012; Millis et al., 2011). Each text focused on one of four research methods topics: *replication, experimenter bias, causality*, and *dependent variables*. Each text had the following components: (1) a title that was designed to inspire text-based interest through uniqueness and novelty (Hidi, 1990), (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. The titles provided little information about the text content in order to avoid the possibility that participants would make selections based on prior knowledge, content familiarity, or potential ease or difficulty (see Figure 1 for all four text titles). 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 title of this text was: "A horse is a horse of course of course, unless that horse is a counting horse." 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 1500 words ($M = 1532$, $SD = 18$) in length.

2.2.2. Text difficulty manipulation

We created an easy and a difficult version of each of the four texts, which resulted in a final set of eight texts (one easy and one difficult text for each topic). Participants received easy versions of the texts for two topics and difficult versions for the remaining two topics. The text difficulty manipulation was achieved by modifying the texts along five dimensions: *sentence length, word frequency, referential cohesion, narrativity,* and *syntactic complexity* (Graesser, McNamara, & Kulikowich, 2011). Easy versions of the texts consisted of shorter, simpler sentences, with few low-frequency words (accomplished by replacing low-frequency words like "scrumptious" with higher-frequency words like "tasty"). The easy texts were also more referentially cohesive (accomplished primarily by replacing ambiguous pronouns with nouns), and were written in a comparatively more narrative or story-like style, rather than an expository style. In contrast, difficult versions of texts consisted of longer, more complex sentences, low-frequency words, lower referential cohesion, and more of an expository style (see Table 1 for examples of the text difficulty manipulations).

Importantly, text length (mean of 1519 and 1527 words for easy and difficult texts, respectively) and content were kept consistent across easy and difficult versions. To objectively verify the text difficulty manipulation, we assessed the linguistic features of the easy and difficult versions of each text using the Flesch–Kincaid Grade Level score (FKGL; Klare, 1974), a widely used measure of the

Table 1
Examples showing manipulations of text difficulty.

Manipulation	Easy	Difficult
Sentence length	Hypotheses are all about relationships between variables. All hypotheses have two parts. The first part involves an independent variable, which is either manipulated by an experimenter or is a participant variable. **(3 short sentences)**	Hypotheses describe relationships between variables, and they have two parts, which are the independent variable, which is either manipulated by an experimenter or is a participant variable. **(1 long sentence)**
Word frequency	Dr. Jones found that people tended to rate the free range chicken as being more **tasty** than the caged chicken.	The results demonstrated that people tended to rate the free range chickens as being more **scrumptious** than the caged chickens.
Referential cohesion	**Dr. Jones** was astounded and confused	**She** was astounded and confused.
Narrativity	Dr. Jones **knew that she needed to provide** an appropriate operational definition for the dependent variable in her study – tastiness.	**She was sentient of the fact that she needed to provide** an appropriate operational definition for the dependent variable in her study – tastiness.
Syntactic complexity	Because of these opinions or biases, this is exactly the sort of experiment where the participants' and the researchers' personal beliefs **could bias the results**.	Because of these opinions or biases that people might have, this is unerringly the sort of experiment in which the participants' and the researchers' personal beliefs **about the possible advantages of free-range chickens over cage-raised chickens could prejudice the results**.

Table 2
Means and standard deviations (in parenthesis) of textual features between easy and difficult texts.

Measure	Easy text	Difficult text	Effect size (*d*)
Flesch–Kincaid Grade Level (FKGL)	9.5 (0.40)	11.1 (0.70)	2.81
Narrativity	62.8 (8.00)	57.6 (4.60)	0.79
Referential cohesion	70.8 (8.70)	64.3 (11.9)	0.62
Deep cohesion	77.3 (12.3)	73.3 (14.1)	0.30
Syntactic simplicity	70.3 (10.4)	54.5 (5.40)	1.90

Note: Higher values indicate increased difficulty for FKGL. Lower values indicate increased difficulty for other measures.

grade-level readability of texts. Flesch–Kincaid scores approximate the level of text difficulty in reference to the grade level of the analyzed text (i.e., a score of 9 indicates that the text is at a ninth grade readability level). The FKGL for easy texts ($M = 9.5$, $SD = 0.4$), was lower than FKGL for difficult texts ($M = 11.1$, $SD = 0.7$) ($d = 2.81$), as was intended. FKGL is somewhat limited in that its computation of text difficulty is based solely on surface-level features, specifically the number of sentences, words, and syllabus. Therefore, we performed a deeper analysis of the text using Coh-Metrix, a computational tool for multidimensional textual analysis (Graesser, McNamara, Louwerse, & Cai, 2004). We focused on the four linguistic features (narrativity, referential cohesion, deep cohesion, and syntactic simplicity) that have been linked to text difficulty (Graesser et al., 2011). These measures are also closely related to the text manipulations discussed above. Table 2 presents descriptive statistics for each feature. Higher values for each of the Coh-Metrix features indicate that the text is easier to read and comprehend. We found effect sizes ranging from 0.30 to 1.90 for the four Coh-Metrix features, with difficult versions having lower values than the easy versions. Taken together, these results confirm that the difficult texts are equally verbose but are more difficult than the easy texts with respect to language and discourse.

2.2.3. Interest-based text preference manipulation

Participants were shown the titles of four texts and were asked to rank the texts based on their interest in each title (similar to Flowerday, Schraw, & Stevens, 2004; Fulmer & Frijters, 2011). The two highest ranked titles were considered *preferred* texts, whereas the remaining two titles were the *non-preferred* texts. Rather than only giving participants their preferred (or non-preferred) text to read, all participants read all four texts. Thus, the interest-based text preference manipulation occurred at the text-level, rather than the activity-level. This within-subjects design eliminated reading-ability and content-related confounds by keeping the overall task and content consistent across students. A McNemar–Bowker test on the relationship between text and ranking was non-significant ($p > .05$), suggesting no statistical relationship between individual texts and how they were ranked.

To reaffirm the manipulation, the following message was displayed before participants read a preferred text: "You said you wanted to read this text first [second]. Congratulations! You get to read an article you wanted to read!". In contrast, participants received this message before reading a non-preferred text: "You said you didn't [definitely didn't] want to read this text. Too bad! Even though you didn't want to read this article, you have to read it anyway."

2.3. Measures

2.3.1. Manipulation checks

We measured participants' perceptions of text difficulty and text interest as manipulation checks for the text difficulty and interest-based text preference manipulations, respectively. Perceived difficulty was assessed with the following question: "How difficult are you finding this text?" Participants responded on a six point scale: (1) very easy, (2) easy, (3) somewhat easy, (4) somewhat difficult, (5) difficult, and (6) very difficult. Perceived interest was measured with the following question: "How interesting are you finding this text?" Participants responded on a similar six-point scale ranging from (1) very uninteresting to (6) very interesting.

2.3.2. Engagement measures

Valence and arousal were measured with the Affect Grid (Russell, Weiss, & Mendelsohn, 1989). The Affect Grid is a single item affect measurement instrument consisting of a 9×9 (valence $\times$ arousal) grid, and is theoretically grounded in the Circumplex model of affect (Russell, 1980, 2003; see Figure 2 for an image of the Affect Grid). It is a validated measure of affect with adequate reliability (Cronbach's alpha = .85), convergent validity (correlations of .89 or higher with similar scales of affect), and discriminant validity (correlations of .08 or lower with dissimilar scales of affect; Russell et al., 1989). The arousal dimension ranges from low arousal (1) to high arousal (9), while the valence dimension ranges from unpleasant feelings (1) to pleasant feelings (9). The Affect Grid was presented on the computer screen and participants selected the cell that best represented their current affective state. The Affect Grid affords the ability to collect self-report affect with a single click, thereby making it particularly convenient for assessing affective engagement at multiple time points with minimal disruption.

Self-paced reading times were computed as the elapsed time between the last and first pages of each text after accounting for time taken to complete the within-text self-reports of valence, arousal, perceived interest, and perceived difficulty.

Mind wandering was measured by prompting participants to indicate if they are thinking of task-unrelated thoughts with auditory thought probes. Prior to the start of the experiment, participants read a definition of mind wandering that was 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 informed that they would hear an auditory tone (i.e., a beep) periodically during reading and they should respond by pressing keys marked "yes" if they were zoning out at the time of the tone or "no" if they were not zoning out. The auditory mind wandering probes were inserted at 10 pseudo-random points throughout each text. Mind wandering was scored as the proportion of "yes" responses to the thought probes.

2.3.3. Learning assessments

To measure learning, we created 48 multiple-choice questions that required deep reasoning according to the Graesser and Person question asking taxonomy (Graesser, Ozuru, & Sullins, 2010; Graesser

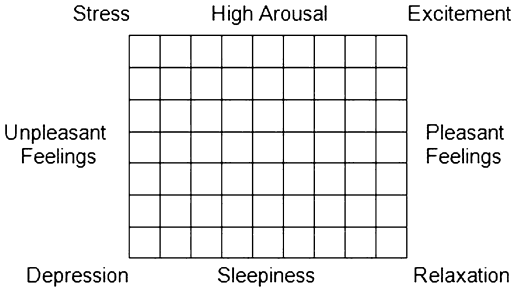

Fig. 2. Affect Grid (based on Russell et al., 1989).

& Person, 1994). There were four questions for each of the three core concepts in each text (12 questions total for each text). Each question had four multiple choice options: (1) the target (the correct response); (2) a near-miss (an option that sounded correct but was not); (3) a thematic miss (an option that followed the theme of the content but was not actually related to the question); and (4) a distractor (an option that was not at all related). There were two 24-item versions of the test that were counterbalanced across participants as pre- and posttests, respectively. Scores for the pretest and the posttest were computed as the proportion of correct responses corresponding to each text.

We also developed a measure that tapped participants' ability to apply the acquired knowledge to new situations in order to demonstrate *transfer*. This test required participants to evaluate the scientific merits of three previously unseen case studies by answering four multiple-choice questions pertaining to each study. There were 12 questions in all, with one question related to each of the three core concepts for each text. The questions required participants to reason whether the case study was flawed or what steps the researcher had taken to avoid potential flaws. The transfer test was scored as the proportion of correct responses to the three questions pertaining to each text.

Example multiple choice and case study questions are listed in the Appendix S1. Both tests were extensively piloted prior to data collection to ensure that the final set of items were not too easy or too difficult in that they could not be answered without reading the texts, but could be answered after reading (data available upon request).

2.3.4. Individual difference covariates

Interest in the overall topic of research methods is expected to correlate with the dependent variables of engagement and learning. Hence, we adapted a topic interest measure from Linnenbrink-Garcia et al. (2010) to measure and covary out students' baseline interest in learning about research methods prior to engaging in the learning task (11 items, Cronbach's alpha = .86). 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 is all about." Responses ranged from 1 (strongly disagree) to 6 (strongly agree). Given the nature of the task, it was also essential to control for individual differences in reading comprehension, which was measured using an abbreviated version of the comprehension subtest from the Nelson–Denny Reading Test (Brown, Fishco, & Hanna, 1993).

2.4. Procedure

Participants were individually tested over a 2–2.5 hour session. The text reading portion of this session lasted on average, 23.4 minutes (SD = 10.5) total for all four texts, so it was not overly cumbersome. The actual study lasted for approximately 1.5 hours. The remaining time after the study was spent completing questionnaires unrelated to the data reported here.

All experimental materials were administered via a computer interface. Figure 3 depicts the various components of the protocol, including key measures. Participants signed an informed consent and then completed the Nelson–Denny Reading Comprehension Test. 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. This was followed by the baseline topic interest in research methods measure and the 24-question multiple choice pretest. Next, participants were instructed on how to use the Affect Grid to self-report their valence and arousal using standard instructions for this measure.

Participants were then shown the four text titles (randomly-ordered for each participant) corresponding to the four topics (replication, experimenter bias, causality, and dependent variables). They were asked to rank order the titles based on their interest in reading the texts (see Figure 1 for a screenshot of the interface). Ranking options were: "want to read first", "want to read second", "don't want to read", and "definitely don't want to read". Their two preferred selections were taken to be the *preferred* texts, while the remaining two were the *non-preferred* texts.

Following the rankings, participants were given instructions about how to respond to the mind-wandering probes. Next, they started the learning session, where they were presented with each of the four texts (order was counterbalanced across participants). Participants were reminded of their interest-based preferences before reading each text. Each text was displayed on the computer screen on a page-by-page basis (approximately 144 words per page). Participants used the space bar to navigate forward but could not return to a previous page. Ten-auditory mind wandering probes were pseudo-randomly inserted at particular pages and participants responded to each probe by selecting specifically labeled "yes" or "no" on the keyboard.

For each text, participants indicated their current affective states via the Affect Grid before, in the middle (after reading half of the total number of pages), and after reading. They self-reported levels of perceived interest and perceived difficulty in the middle and after reading each text.

After reading all four texts, participants completed the 24-item learning assessment and the 12-item transfer test. Participants were then debriefed and compensated.

2.5. Data treatment

Each of the 84 participants read four texts, which yielded 336 (84×4) cases. There were 10 cases where participants spent less than a minute reading and completing the within-text measures. For these 10 cases, participants were deemed unresponsive to the instructions, and these cases were removed prior to data analysis. To ensure that removal of these cases did not unduly influence the results, all analyses were repeated without removal of any cases. All significant effects and interactions when including all cases remained significant at $p < .05$ after removing the 10 cases. One effect that was marginally significant when including all cases became statistically significant at $p < .05$ after removing the 10 cases. This was the effect of interest-based text preference on valence ($B = .183$, $SE = .116$, $p = .08$ in the all cases model; $B = .217$, $SE = .118$, $p < .05$ after removing the 10 cases). In order to account for inter-text differences, all dependent variables were standardized at the text level. An examination of the histograms of the individual variables did not yield any significant departures from normality, nor were there any outliers. Skewness and kurtosis values for all dependent variables are reported in Table 3 and were mostly within the –1/+1 recommended range.

Table 3
Skewness and kurtosis values for dependent variables.

Measure	Skewness	Kurtosis
Interest	.14	–1.25
Difficulty	.53	–.49
Valence	–.27	–.74
Arousal	.30	–1.13
Reading time	.89	.64
Mind wandering	.42	–1.04
Posttest	–.07	–.73
Transfer	.35	–1.02

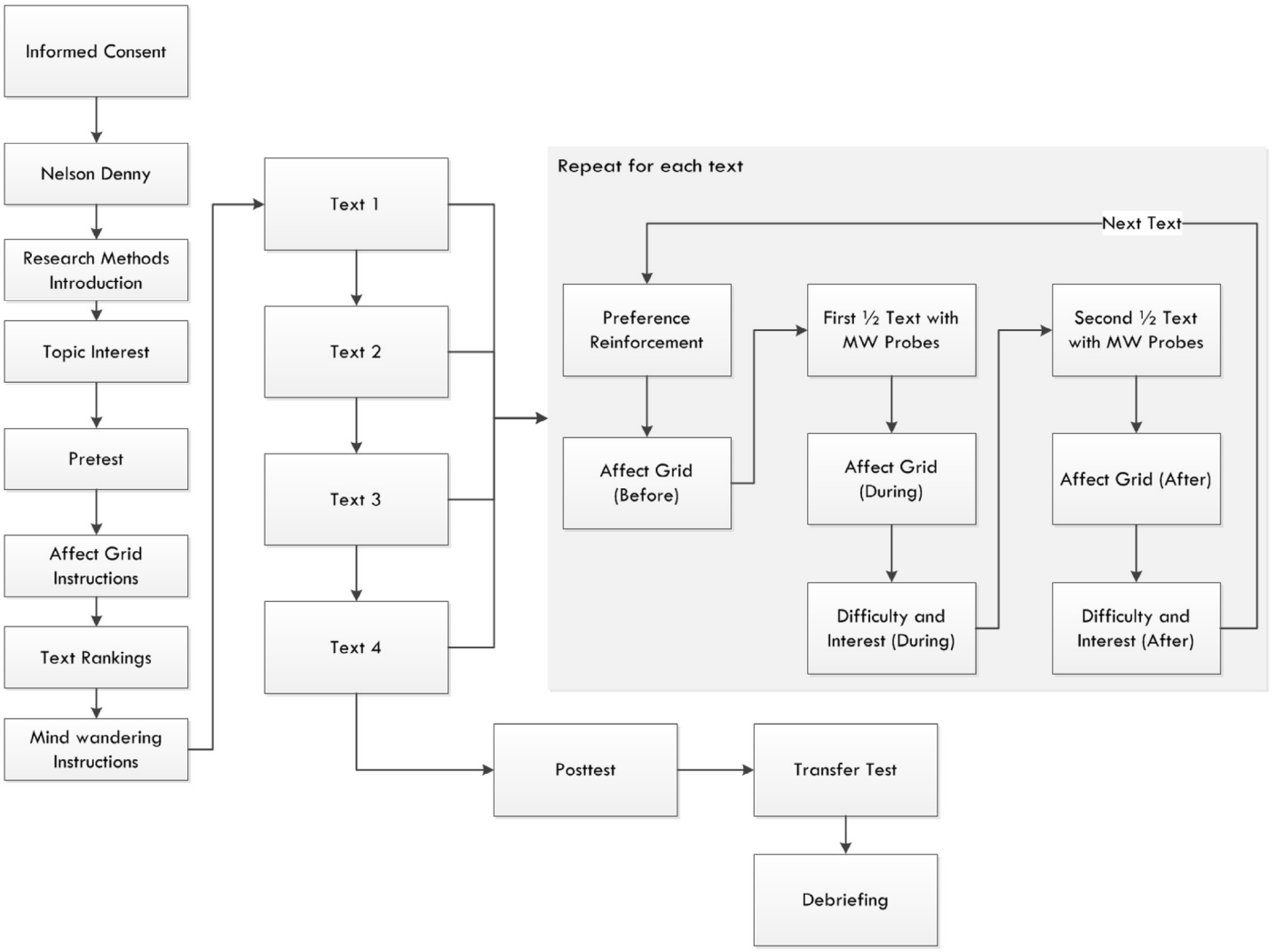

Fig. 3. Experimental protocol.

3. Results

Data were analyzed at the text-level. Unstandardized descriptive statistics for main effects and the interest-based text preference × text difficulty interaction are shown in Table 4. Due to the repeated and nested structure of the data, a mixed-effects modeling approach was adopted for all analyses (Pinheiro & Bates, 2000). Mixed-effects models include a combination of fixed and random effects and can be used to assess the influence of the fixed effects on dependent variables after accounting for any extraneous random effects. The *lme4* package in R (Bates & Maechler, 2010) was used to perform the requisite computation. The random effect

Table 4
Raw (unstandardized and not corrected for covariates) means and standard deviations (in parentheses).

Measures	Main effects				Text difficulty × interest-based preference interaction			
	Interest-based preference		Text difficulty		Non-preferred		Preferred	
	Non-preferred	Preferred	Easy	Difficult	Easy	Difficult	Easy	Difficult
Manipulation checks								
Avg. interest [1–6]	**2.01 (1.16)**	**2.22 (1.12)**	2.19 (1.07)	2.06 (1.14)	2.08 (1.28)	2.01 (1.38)	2.33 (1.37)	2.12 (1.45)
Avg. difficulty [1–6]	1.45 (1.13)	1.30 (1.06)	**1.24 (1.01)**	**1.47 (1.13)**	1.32 (1.13)	1.51 (1.30)	1.15 (1.08)	1.40 (1.19)
Engagement								
Avg. arousal [1–9]	4.02 (1.85)	4.07 (1.81)	4.07 (1.75)	4.01 (1.83)	3.91 (2.09)	4.05 (2.10)	4.22 (2.01)	4.03 (2.17)
Avg. valence [1–9]	**4.40 (1.46)**	**4.69 (1.59)**	4.48 (1.57)	4.66 (1.45)	4.41 (1.81)	4.44 (1.58)	4.60 (1.72)	4.87 (1.73)
Read. time (s) [0]	379 (150)	374 (174)	**364 (154)**	**393 (164)**	373 (160)	392 (165)	**360 (174)**	**400 (188)**
Prop. MW [0–1]	0.27 (0.27)	0.27 (0.27)	0.29 (0.27)	0.25 (0.25)	**0.32 (0.33)**	**0.25 (0.29)**	0.28 (0.28)	0.28 (0.30)
Learning								
Prop. post [0–1]	0.37 (0.17)	0.40 (0.18)	0.38 (0.20)	0.39 (0.16)	0.37 (0.23)	0.37 (0.20)	0.40 (0.23)	0.41 (0.20)
Prop. transfer [0–1]	0.26 (0.18)	0.25 (0.17)	0.24 (0.17)	0.29 (0.19)	0.28 (0.24)	0.26 (0.24)	**0.22 (0.22)**	**0.31 (0.25)**

Notes: Possible range of measures shown in square brackets. Avg. is the average of the measure across the during and after time points. MW = mind wandering; Read. = reading; Prop. = proportion. Significant effects bolded.

Table 5
Model coefficients (B) for mixed effects models.

Effect	Avg. interest	Avg. difficulty	Avg. valence	Avg. arousal	Reading time	Prop. mind wandering	Prop. posttest	Prop. transfer
Intercept	–1.040	–0.598	–0.792	0.116	0.491	0.179	–0.556	0.388
Basic covariates								
Compensation credit (vs. cash)	**–0.347**	0.397	**–0.166**	–0.163	–0.454	0.108	–0.247	–0.165
Order 2 (vs. Order 1)	0.024	–0.084	–0.087	–0.069	**–0.288**	–0.020	0.059	–0.017
Order 3 (vs. Order 1)	0.049	0.027	–0.158	0.123	**–0.565**	0.145	0.207	–0.015
Order 4 (vs. Order 1)	0.047	–0.177	–0.152	0.025	**–0.632**	0.041	0.031	–0.342
Individual difference covariates								
Reading comprehension	0.148	**–0.964**	–0.320	–0.188	0.776	–0.226	**1.064**	**0.716**
Baseline topic interest	**0.268**	0.171	**0.249**	–0.004	–0.020	0.006	0.007	–0.116
Model-specific covariates								
Valence before			**0.379**					
Arousal before				**0.479**				
Pretest							**0.130**	0.035
Factors								
Difficult (vs. easy)	–0.007	**0.214**	0.070	0.150	**–0.011**	–0.397	0.155	–0.062
Preferred (vs. non-preferred)	**0.237**	–0.109	**0.217**	0.122	–0.171	–0.274	0.185	–0.227
Difficulty × preference	–0.132	0.006	–0.107	–0.074	**0.234**	**0.469**	–0.084	**0.440**[a]

Note: Bolded values indicate coefficients are significant at $p < .05$.
[a] $p = .051$. All dependent variables are *z*-score standardized by text. Avg. is the average of the measure across the during and after time points. Prop. = proportion.

was *participant*. The fixed effects were *interest-based text preference* (preferred vs. non-preferred), *text difficulty* (easy vs. difficult), and the *interest-based text preference* × *text difficulty* interaction. Reading comprehension and baseline topic interest in research methods were added as fixed effects in all models to control for these individual differences. *Compensation* (course credit or monetary) was also included as a categorical fixed effect in order to control for any effect of compensation (9 out of the 84 participants were monetarily compensated). Finally, text presentation *order* (counterbalanced across participants) was included as a four-level categorical fixed effect in order to account for ordering and fatigue effects. To further account for possible order effects, the three-way order × text difficulty × interest-based text preference interaction, as well as the two-way interactions between order and both text difficulty and interest-based text preference, were included as predictors in all models. Order did not significantly interact with either difficulty or preference, and the three-way interaction was not significant in any of the models. For parsimony, these two- and three-way interactions with order were not included in the final models reported in the Results.

Table 5 summarizes the model coefficients for all predictors, including the covariates, for each of the mixed-effects models. A two-tailed significance threshold of .05 was adopted for all analyses. Degrees of freedom vary slightly across analyses due to occasional missing data. Although the central focus of our analyses is the interaction between interest-based text preference and text difficulty, we report all significant direct effects and interactive effects. We do so in the interest of full disclosure of our analyses and results and to provide a comprehensive report of these analyses.

3.1. Manipulation checks

We first considered the influence of interest-based text preference and text difficulty on participants' perceptions of text interest and difficulty. Interest during and after reading were strongly correlated ($r = .79$, $p < .001$), so we proceeded by averaging the interest ratings at these two time points. The model for average perceived interest was significant, $F(1, 299) = 4.34$, $p = .038$. Participants reported higher levels of interest for preferred texts compared with the non-preferred texts ($B = .237$, $SE = .124$). There was no significant effect of text difficulty or the interest-based text preference × text difficulty interaction for average perceived interest.

Perceived difficulty during and after reading were also strongly correlated ($r = .89$, $p < .001$), so we averaged the two. The model for average perceived difficulty yielded a significant text difficulty effect, $F(1, 300) = 11.0$, $p = .001$, but no significant effect of interest-based text preference or the interaction. These findings confirmed that average perceived difficulty was higher for the difficult texts compared with the easy texts ($B = .215$, $SE = .106$).

Perceived interest and perceived difficulty were negatively correlated ($r = –.34$, $p < .05$). This suggests that interest-based text preference might reduce the perception of text difficulty, but interest-based text preference did not emerge as a significant predictor of average perceived difficulty. However, when perceived difficulty during and after reading were examined individually, there was a significant effect of interest-based text preference on perceived difficulty after reading, $F(1, 299) = 4.34$, $p = .04$, but not during reading. Thus, participants reported lower levels of difficulty after reading the preferred texts compared with the non-preferred texts ($B = –.145$, $SE = .113$), even though the objective (text) difficulty was equivalent.

3.2. The effects of interest-based text preference and text difficulty on engagement

The during and after reading measurements of valence ($r = .85$, $p < .001$) and the during and after reading measurements of arousal ($r = .86$, $p < .001$) were strongly correlated, so we proceeded by averaging across the two time points. Pre-reading levels of valence and arousal were also included as covariates in the respective models in order to control for baseline effects. Interest-based text preference was a significant predictor of average valence $F(1, 298) = 4.16$, $p = .042$, with higher levels of valence for the preferred texts compared with the non-preferred texts ($B = .217$, $SE = .118$). Interest-based text preference was not a significant predictor of average arousal. There was no significant text difficulty main effect or interest-based text preference × text difficulty interaction on average arousal or average valence.

The model for reading times did not yield a significant effect for interest-based text preference, indicating that there were no significant differences in reading times between preferred texts versus non-preferred texts. There was a significant text difficulty effect, $F(1, 300) = 5.32$, $p = .02$ and a significant interest-based text preference × text difficulty interaction, $F(1, 30) = 3.95$, $p = .047$. We focus on the interaction in lieu of the main effect of difficulty, as the significant interaction suggests that the main effect may be misleading (Maxwell & Delaney, 2004). A post-hoc test indicated that there was no significant text difficulty effect for the non-preferred texts, but there was a text difficulty effect for the preferred texts, $t(300) = 2.96$,

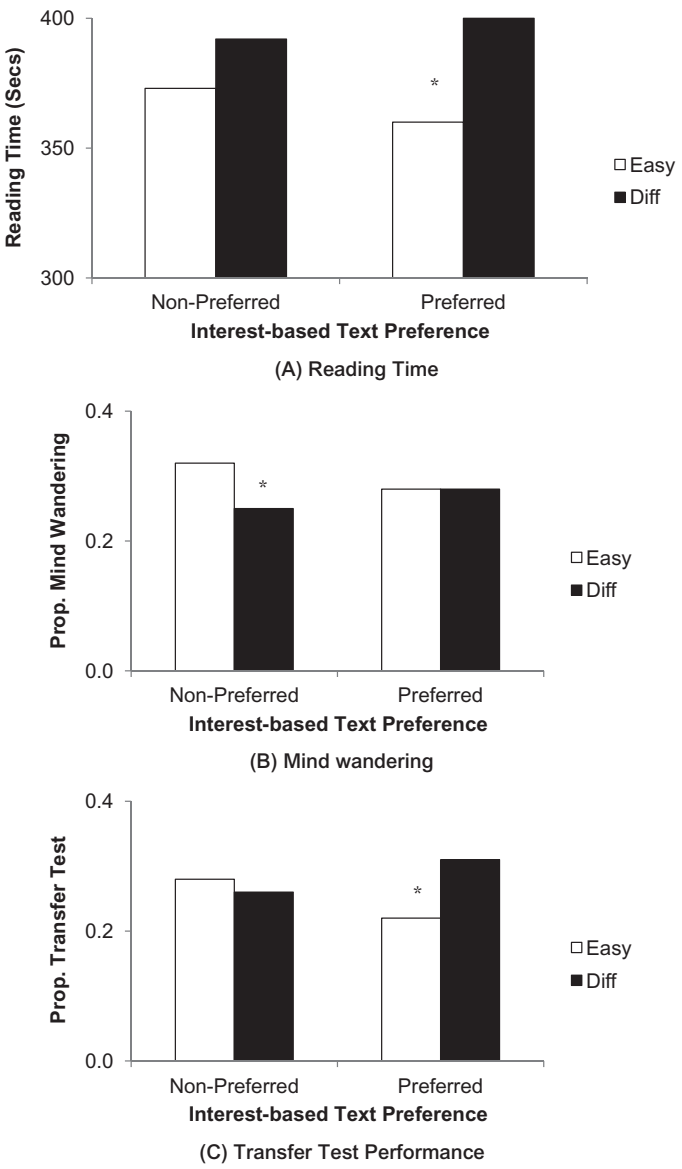

Fig. 4. Significant (* $p < .05$) interactions between interest-based text preference and text difficulty.

$p = .01$ (see Figure 4A). Participants spent more time reading preferred difficult texts compared with the preferred easy texts ($B = .171$, $SE = .075$).

Neither interest-based text preference nor text difficulty emerged as significant predictors of mind wandering, but the interest-based text preference × text difficulty interaction was significant, $F(1, 256) = 4.87$, $p = .03$. Post-hoc tests indicated that text difficulty had no significant effect on mind wandering for the preferred texts. However, for non-preferred texts, there was significantly less mind wandering while reading the difficult texts ($B = -.40$, $SE = .140$) compared with the easy texts, $t(256) = 2.84$, $p = .01$ (see Figure 4B).

3.3. The effects of interest-based text preference and text difficulty on learning

The models for learning outcomes included pretest scores as an additional covariate in order to control for prior knowledge. The model for posttest scores did not yield significant effects for interest-based text preference, text difficulty, or their interaction. However, the model for transfer test performance yielded a marginally significant interest-based text preference × text difficulty interaction, $F(1, 299) = 3.96$, $p = .051$. Post-hoc tests indicated that text difficulty had no significant effect on transfer test performance for the non-preferred texts. However, scores on the transfer test were significantly higher for the difficult compared with the easy texts ($B = .378$, $SE = .156$), but only when they were preferred, $t(299) = 2.42$, $p = .02$ (see Figure 4C). This effect was also observed after controlling for performance on the posttest, $t(298) = 2.44$, $p = .015$, $B = .382$, $SE = .156$, which suggests that the difficulty effect on transfer performance for the interest-based preferred texts was the net of posttest performance.

4. Discussion

Considerable theory in educational and cognitive psychology has focused on the benefits of moderate difficulty in promoting engagement and learning (e.g., Bjork & Bjork, 2011; Deci & Ryan, 1985). However, the empirical research on the effects of difficulty is somewhat conflicting, suggesting that higher levels of difficulty can be either beneficial or detrimental to engagement and learning depending on the context and individual. There is a complementary line of work on the benefits of autonomy and interest in supporting engagement and learning, but this is generally independent of difficulty. Our work connects these two lines of research by testing the role of interest-based text preference as a moderator of the effect of text difficulty on engagement and learning.

To test this hypothesis, participants studied four instructional texts on research methods topics in a 2 × 2 interest-based text preference (preferred vs. non-preferred texts) × text difficulty (easy vs. difficult) within-subjects experiment. We selected research methods topics because our participants (psychology undergraduates) would have some interest in the subject, but would also experience some degree of difficulty as they learned more about the intricacies of research methods. Additionally, research methods are often associated with negative affective states (i.e., anxiety, frustration) and feelings of low self-efficacy (Murtonen & Lehtinen, 2003; Papanastasiou & Zembylas, 2008). Because this topic is an essential element to undergraduate social science programs, it is important to investigate how we can alter aspects of the learning context in order to increase engagement while learning research methods. In the remainder of this section, we consider the extent to which the data support this hypothesis, followed by a discussion of the limitations of the present study and suggestions for future research.

4.1. The moderating role of interest-based text preference

The current study experimentally manipulated text difficulty (easy vs. difficult) and interest-based text preference (preferred vs. non-preferred texts), and examined their interactive effects on six outcomes: valence, arousal, mind wandering, reading time, learning (posttest), and knowledge transfer. We had two specific hypotheses regarding the moderating role of interest-based text preference: (1) reading preferred texts would result in similar levels of engagement and learning across difficult and easy texts; or (2) reading preferred texts would have an added benefit, supporting higher engagement and learning for the difficult compared with easy texts (Fulmer & Frijters, 2011). Interest-based text preference interacted with text difficulty to predict reading times, mind wandering, and knowledge transfer, but the interaction did not predict valence, arousal, or posttest scores.

In support of the first hypothesis, participants' reports of mind wandering during preferred texts were similar regardless of the level of text difficulty. However, for the non-preferred texts, participants reported more mind wandering for the easy texts compared with the difficult texts (contradictory to Feng et al., 2013).

One possible explanation is that participants were not investing enough effort while reading a non-preferred easy text, so they had more resources to mind wander. This is consistent with the executive resources view of mind wandering, which posits that mind wandering is more likely for easy tasks, as there are more available resources to entertain task-unrelated thoughts because easy tasks do not fully sustain executive resources (Feng et al., 2013; Smallwood & Schooler, 2006). Participants' reports of mind wandering during preferred texts were similar regardless of the level of text difficulty, indicating that reading preferred texts mitigated the influence of text difficulty on mind wandering. Hence, the present study also puts a boundary condition on the relationship between mind wandering and text difficulty.

In support of the second hypothesis, participants spent more time reading difficult texts, but only in the preferred text condition. Feng et al. (2013) found that learners spent more time reading a difficult text compared with an easy text. Our results qualify their finding by showing that learners appropriately calibrate their resources by increasing reading time when text difficulty increases (Feng et al., 2013; Vega et al., 2013), but only when the text is preferred. These findings also support the theoretical position that higher difficulty leads to increased effort and persistence (Clifford, 1990), but suggest that this is more likely to occur when the learner expresses a preference for the text.

Also supporting the second hypothesis, knowledge transfer scores were higher for preferred difficult texts, compared with preferred, easy texts. There was no difference in transfer test performance related to text difficulty for non-preferred texts. The difficulty of the current texts, which involved less frequently used terms, more complex syntax, and less cohesion, can be regarded as a form of desirable difficulties that require more effort and deeper levels of processing for appropriate comprehension (Bjork & Bjork, 2011; McNamara & Kintsch, 1996). Thus, it is possible that the difficult texts facilitated the higher-level inference needed for the transfer test through encouraging deeper levels of processing. However, the beneficial effects of the difficult texts were only realized when learners invested resources by calibrating their effort to match the level of difficulty. Thus, our findings suggest that interest-based text preference increased participants' effort on difficult texts (shown to some extent through increased reading time), and this greater effort may have contributed to deeper processing and, consequently, improved knowledge transfer.

It is worth mentioning that the interaction between interest-based text preference and text difficulty predicted performance on the transfer test, but not on the posttest. This is likely due to the fact that these two learning assessments tapped different knowledge components and differed in difficulty. In particular, although both learning assessments required some inference, the transfer test demanded higher-level inferences compared with the posttest, making it more difficult than the posttest. Evidence for differences in difficulty comes from the significantly lower scores on the transfer test ($M = .258$, $p = .131$) compared with the posttest ($M = .385$, $SD = .146$), paired-samples $t(82) = 6.26$, $p < .001$. The fact that the interactive effect on transfer performance replicated after controlling for posttest scores provides some evidence that these tests target different levels of knowledge. Thus, it might be possible that interest-based text preference in combination with higher text difficulty inspired the level of deep processing needed for success on the transfer test, but not necessarily for the posttest.

Overall, reading interest-based preferred texts was an important motivational resource for supporting engagement and learning (Hidi, 1990), and its effectiveness was most apparent during tasks of higher difficulty. When engaging in difficult tasks, learners may require additional incentives to calibrate their effort, remain attentive, and sustain their engagement to foster learning. Expressing one's interest-based preferences in texts appears to be one such incentive that both maintains engagement during difficult texts as well as inspires deeper processing, thereby leading to greater transfer. Further evidence for the value of preferred texts in the context of higher difficulty comes from the fact that the preferred texts were perceived as less difficult than non-preferred texts after reading the entire text (Durik & Matarazzo, 2009; Renninger et al., 2002), despite these difficult texts being equivalent on objective difficulty.

Although our emphasis was on the interactive effect of these two factors, we tested their direct effects as well. Out of the 12 possible direct effects, we found only one main effect of interest-based text preference and no meaningful direct effect[1] for text difficulty. The sole direct effect indicated that reading preferred texts was related to more positive valence. This finding is not surprising, as interest is considered by some to be a positively-valenced emotion (Ainley et al., 2005; Izard, 2009), and higher interest is consistently linked to reports of positive valence during learning (Ainley, 2012; Pekrun et al., 2002). This was shown in the present data via the higher perceptions of interest when reading preferred texts compared with non-preferred texts (our manipulation check) and that perceived interest was a positive predictor of valence[2]. It is important to note that this positive impact of interest-based text preference occurred regardless of the level of text difficulty, as we did not find a significant interaction between interest-based text preference and text difficulty for valence.

We also did not find any direct or interactive effect of interest-based text preference or text difficulty on arousal. It may be that reading preferred versus non-preferred texts has no impact on the arousal dimension. However, it also may be that reading preferred versus non-preferred texts increases or decreases arousal depending on the individual. For example, reading preferred texts may lead to positive valence, but this positive valence could involve feelings of alertness/excitement (high arousal) or feelings of calmness/relief (low arousal). Additional affective measurements are needed in order to consider these possibilities.

4.2. Limitations and further research

It is important to acknowledge some limitations with the present study. The measurement of engagement continues to be a contested domain (Shernoff, 2013). We incorporated both self-report and behavioral measures of engagement, which offered a more comprehensive picture of engagement during learning from instructional texts. However, we recognize that there are some limitations to these approaches. The first set of limitations pertains to our measurement of engagement. Some of the self-reported measures, including perceived difficulty and perceived interest, were assessed with a single item. Although single item measures provide less overall information compared with lengthier scales, recent studies suggest that single-item measures can have acceptable reliability and validity, particularly if the purpose of the measure is clear, the item accurately represents the underlying construct, and the assessment is contextually-grounded and occurs in the moment (Ainley et al., 2005; Gogel et al., 2014). One issue with our mind wandering measure was that participants were periodically interrupted by the mind wandering probes, which might have interfered with comprehension and learning processes. Furthermore, the mind wandering measure focused solely on task-unrelated thoughts and

[1] There was a significant conditional main effect of text difficulty on reading time; however, there was also a significant interaction between interest-based text preference and text difficulty on reading time. The interaction suggests that the significant main effect of text difficulty on reading time holds true only for preferred texts.

[2] A mixed effects model indicated that average perceived interest positively predicted average valence ($F(1, 300) = 62.4$, $p < .001$, $B = .301$, $SE = .050$) after controlling for valence before reading, baseline topic interest, reading comprehension, compensation, and text order.

did not attempt to categorize these thoughts more precisely, a limitation that should be addressed with categorical measures of mind wandering (Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011). Similarly, other than the influence of text difficulty, we did not investigate the many potential reasons why participants spent more or less time reading a particular text. This information could be collected via think-aloud protocols, but this has the added potential to interrupt the primary learning activity. Together, these limitations point to the challenge of balancing the information depth and precision vs. the intrusiveness of online measurement of engagement. More basic research and methodological advances are needed to resolve these issues.

A second limitation is that participants were not given any explicit reasons for studying the instructional texts. This was not an omission but an intended decision in order to alleviate concerns that providing participants with explicit goals for reading (e.g., reading to search for particular content) might increase variability by eliciting different reading strategies (McCrudden & Schraw, 2007). That being said, it is possible that the lack of explicit instructions might have led participants to implicitly adopt different strategies based on how they contextualized the learning task. Hence, it would be informative to replicate the current study with explicit reading goals to ensure that participants are on equal footing. The third limitation was related to our manipulation of interest-based text preference. This manipulation asked participants to rank four texts based on how much they wanted to read each text. Participants' rankings do not provide us with specific information about the degree of their preferences for one text over another. For some participants, there may have been only a marginal difference in preference between the second ("preferred") and third ("non-preferred") choice, while others may have perceived a larger difference in preference. We are also unable to separate the effects of the rankings from the effects of the reinforcing statements when participants were presented with each text. Additionally, participants' rankings were based only on the titles of the text. We acknowledge that participants could make more accurate interest-based rankings had they read a brief summary of the content of the text prior to ranking the texts. Future research may consider asking participants to rate their level of preference for each text based on a brief summary or an excerpt of the contents.

The fourth limitation was that our manipulation of text difficulty was not sensitive to individual differences in reading ability and prior knowledge. In the present study, the level of difficulty was manipulated at the text level. As a result, the level of difficulty was independent of participants' levels of ability in both reading comprehension and the area of research methods. Although we controlled for participants' general abilities in reading comprehension, future research should continue to focus on the relationship between ability and difficulty. One solution to this issue is to prepare texts with multiple levels of difficulty and then assign texts to participants in a manner that is sensitive to their levels of ability (Fulmer & Tulis, 2013). Similarly, an important next step is to consider the curvilinear relationship between text difficulty and engagement by considering easy, moderate, and difficult texts, as opposed to the two levels of difficulty considered in this study. This would contribute to our understanding of how task characteristics interact with learner characteristics to predict engagement, and how to best calibrate the task to the individual learner.

Future research should consider the causal pathways among task characteristics, engagement, and learning as individuals engage in learning activities. For example, it may be that interest-based text preferences increased participants' valence, which then heightened cognitive and behavioral engagement, subsequently resulting in deeper learning. Future research should also examine how difficulty interacts with other aspects of the task, learning environment, and individual to predict engagement and learning. This includes observing these relationships with texts of a variety of topics, such as those that may be more intrinsically interesting for students. This also includes investigating how other motivational strategies, such as the provision of choice or the support of task value, interact with difficulty to influence engagement and learning (Patall, Sylvester, & Han, 2014). Finally, because this was a laboratory study with a sample of 84 participants, replication in more authentic educational contexts is needed.

4.3. Concluding remarks

Learners' engagement and performance are influenced by the conditions of the learning activity. Thus, it is important that research continues to focus on how various features of the learning task foster engagement and enhance learning. This is particularly true for areas of study that are related to low engagement due to students' negative perceptions and affect toward the topic (i.e., research methods). Research is also needed to resolve the controversial findings regarding the effects of difficulty on engagement and learning. The present study contributed to this research by considering how text difficulty and the opportunity for learners to express their preferences based on interest influenced engagement and learning when studying texts on research methods topics. Contributing to the contradictory findings regarding the effects of difficulty, we found zero significant direct effects of text difficulty on engagement and learning after accounting for the interaction between text difficulty and interest-based text preference. However, we found partial support for the hypothesis that interest-based text preference would moderate the effect of text difficulty on engagement and learning. In particular, reading texts that one had indicated as preferred (based on interest) served as a motivational resource (Hidi, 1990) for increasing reading time, reducing mind wandering, and fostering knowledge transfer while reading difficult instructional texts. Accordingly, the present study contributes to our understanding of the conditions under which difficulty can have beneficial effects on engagement and learning, suggesting that interest-based text preference can be valuable in the context of reading difficult instructional texts. Thus, when engaging in difficult tasks, it may be useful to provide learners with opportunities to express their autonomy (e.g., through expressing preferences) in a manner that aligns with their interests. The fact that the present study was conducted in a laboratory environment limits direct implications in the classroom, at least until it is replicated in more ecological contexts. However, the present results do suggest that there is something important about the opportunity to express preferences during difficult tasks, a context that is all too familiar to learners of any age.

Appendix: Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.cedpsych.2014.12.005.

References

Acee, T. W., Kim, H., Kim, H., Kim, J., Chu, H.-N. R., Kim, M., et al. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, *35*, 17–27.

Ainley, M. (2006). Connecting with learning: Motivation, affect and cognition in interest processes. *Educational Psychology Review*, *18*, 391–405.

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 283–302). New York: Springer.

Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction*, *15*, 433–447.

Ainley, M., & Hidi, S. (2014). Interest and enjoyment. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 205–227). New York: Routledge.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*, 545–561.

Alexander, P. A., & Jetton, T. L. (1996). The role of importance and interest in the processing of text. *Educational Psychology Review*, *8*, 89–121.

Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, *22*(6), 437–442.

Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review*, *64*, 359–372.

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*, 223–241.

Bates, D. M., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes. Available from http://CRAN.R-project.org/package=lme4.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology*, *45*, 180–191.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.

Brown, J. A., Fishco, V. V., & Hanna, G. (1993). G. Forms & H. Rolling Meadows (Eds.), *Nelson–Denny reading test: Manual for scoring and interpretation*. Itasca, IL: Riverside Publishing.

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.), (2012). *Handbook of research on student engagement*. New York: Springer.

Clifford, M. (1990). Students need challenge, not easy success. *Educational Leadership*, *48*, 22–26.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York: Harper-Perennial.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self determination in human behavior*. New York: Plenum Press.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*, 325–346.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and italicised): Effects of disfluency on educational outcomes. *Cognition*, *118*, 111–115.

Durik, A. M., & Matarazzo, K. L. (2009). Revved up or turned off? How domain knowledge changes the relationship between perceived task complexity and task interest. *Learning & Individual Differences*, *19*, 155–159.

Efklides, A. (2002). Feelings and judgments as subjective evaluations of cognitive processing: How reliable are they? *Psychology: The Journal of the Hellenic Psychological Society*, *9*, 163–184.

Efklides, A., & Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction*, *15*, 415–431.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*, 586–592.

Flowerday, T., Schraw, G., & Stevens, J. (2004). The role of choice and interest in reader engagement. *The Journal of Experimental Education*, *72*, 93–114.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59–109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). New York: Springer.

Fulmer, S. M., & Frijters, J. C. (2011). Motivation during an excessively challenging reading task: The buffering role of relative topic interest. *The Journal of Experimental Education*, *79*, 185–208.

Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction*, *27*, 11–20.

Giambra, L. M. (1995). A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness & Cognition*, *4*, 1–21.

Gogel, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., et al. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*, 188–205.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*(2), 371–398.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods. Instruments, and Computers*, *36*, 193–202.

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. McKeown & G. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). New York: Guilford.

Graesser, A., & Person, N. (1994). Question asking during tutoring. *American Education Research Journal*, *31*, 104–137.

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, *48*, 9–26.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology. General*, *114*(3), 357.

Halpern, D. F., Millis, K., Graesser, A., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, *7*, 93–100.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*, 549–571.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*, 111–127.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology*, *60*, 1–25.

Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in Psychological Science*, *21*(5), 348–354.

Klare, G. (1974). Assessing readability. *Reading Research Quarterly*, *10*, 62–102.

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*, 383–409.

Li, W., Lee, A., & Solmon, M. (2007). The role of perceptions of task difficulty in relation to self-perceptions of ability, intrinsic value, attainment value, and performance. *European Physical Education Review*, *3*, 301–318.

Linn, M. C., Chang, H., Chiu, J., Zhang, Z., & McElhaney, K. (2011). Can desirable difficulties overcome deceptive clarity in scientific visualizations? In A. Benjamin (Ed.), *Successful remembering and successful forgetting: a Festschrift in honor of Robert A. Bjork* (pp. 235–258). New York: Psychology Press.

Linnenbrink, E. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation and engagement. In P. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp. 107–124). San Diego, CA: Academic Press.

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., et al. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement*, *70*(4), 647–671.

Linnenbrink-Garcia, L., Patall, E. A., & Messersmith, E. E. (2013). Antecedents and consequences of situational interest. *British Journal of Educational Psychology*, *83*(4), 591–614.

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1–3.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*(2), 113–139.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288.

McVay, J., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology*, *141*, 302–320.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp. 169–196). London: Springer-Verlag.

Moneta, G. B., & Csikszentmihalyi, M. (1996). The effect of perceived challenges and skills on the quality of subjective experience. *Journal of Personality*, *64*, 274–310.

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 11–18.

Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education*, *28*, 171–185.

Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.

Papanastasiou, E. C., & Zembylas, M. (2008). Anxiety in undergraduate research methods courses: Its nature and implications. *International Journal of Research and Method in Education*, *31*, 155–167.

Patall, E. A., Sylvester, B. J., & Han, C.-W. (2014). The role of competence in the effects of choice on motivation. *Journal of Experimental Social Psychology*, *50*, 27–44.

Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R. H., & Perry, R. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). New York: Springer.

Pekrun, R., & Schutz, P. A. (2007). Where do we go from here? Implications and future directions for inquiry on emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 313–331). Amsterdam: Elsevier.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.

Pintrich, P. R., & De Groot, E. (1990). Motivated and self-regulated learning components of academic performance. *Journal of Educational Psychology*, *82*, 33–40.
Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, *98*, 209–218.
Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*, 257–267.
Renninger, K. A., Ewen, L., & Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. *Learning and Instruction*, *12*, 467–491.
Reschly, A., & Christenson, S. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–19). Berlin: Springer.
Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.
Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.
Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.
Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*, 299–324.
Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out while reading: Evidence for dissociations between experience and metaconsciousness. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 203–226). Cambridge, MA: MIT Press.
Seibert, P. S., & Ellis, H. C. (1991). Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory and Cognition*, *19*, 507–513.
Shernoff, D. J. (2013). Measuring student engagement in high school classrooms and what we have learned. In D. Shernoff (Ed.), *Optimal learning environments to promote student engagement* (pp. 77–96). New York: Springer.
Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*, 158–176.
Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13–35). New York: Routledge.
Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14*, 230–236.
Smallwood, J. M., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946–958.
Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381.
Vega, B., Feng, S., Lehman, B., Graesser, A., & D'Mello, S. (2013). Reading into the text: Investigating the influence of text complexity on cognitive engagement. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th international conference on educational data mining* (EDM 2013) (pp. 296–299). International Educational Data Mining Society.
Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91–122). San Diego, CA: Academic.
Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *21*, 3–17.