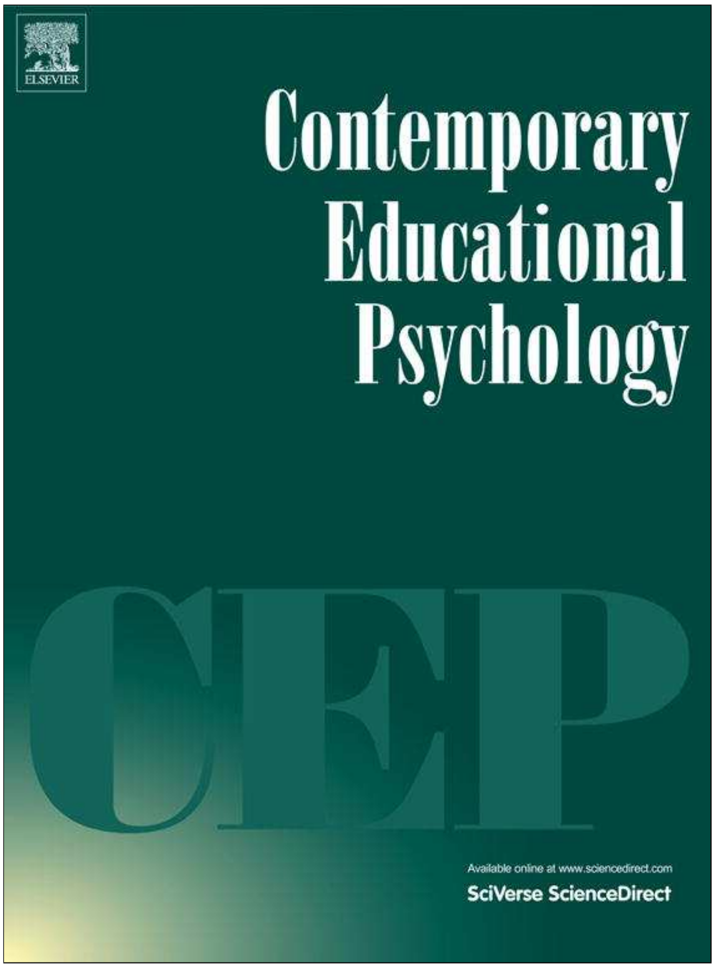


(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Contemporary Educational Psychology 38 (2013) 22–39

Contents lists available at SciVerse ScienceDirect

Contemporary Educational Psychology

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

Using a false biofeedback methodology to explore relationships between learners' affect, metacognition, and performance

Amber Chauncey Strain [a,*], Roger Azevedo [b], Sidney K. D'Mello [c]

[a] University of Memphis, United States
[b] McGill University, Canada
[c] University of Notre Dame, United States

A R T I C L E I N F O

Article history:
Available online 11 August 2012

Keywords:
Affect
Metacognition
Learning outcomes
Self-regulated learning

A B S T R A C T

We used a false-biofeedback methodology to manipulate physiological arousal in order to induce affective states that would influence learners' metacognitive judgments and learning performance. False-biofeedback is a method used to induce physiological arousal (and resultant affective states) by presenting learners with audio stimuli of false heart beats. Learners were presented with accelerated, baseline, or no heart beat (control) while they completed a challenging learning task. We tested four hypotheses about the effect of false-biofeedback. The *alarm* vs. *alert hypothesis* predicted that false biofeedback would be appraised as either a signal of distress and would impair learning (alarm), or as a signal of engagement and would facilitate learning (alert). The *differential biofeedback hypothesis* predicted that the alarm and alert effects would be dependent on the type of biofeedback (accelerated vs. baseline). The *question depth hypothesis* predicted that these effects would be more pronounced for challenging inference questions. Lastly, the *self* vs. *recording hypothesis* predicted that effects would only occur if participants believed that false biofeedback was indicative of their own physiological arousal. In general, learners experienced more positive/activating affective states, made more confident metacognitive judgments, and achieved higher learning when they received accelerated or baseline biofeedback while answering a challenging inference question, irrespective of the perceived source of the biofeedback. Thus, our findings supported the alert and question depth hypotheses, but not the differential biofeedback or self vs. recording hypotheses. Implications of the findings for the integration of affective processes into models of cognitive and metacognitive processes during learning are discussed.

1. Introduction

Beginning in middle school and continuing through high school and beyond, students have to learn about difficult and conceptually-rich topics in mathematics, physics, ecology, chemistry, and biology. It is in these domains that adolescents and young adults face the greatest challenges to learning (PISA, 2009) because they are confronted with novel and unfamiliar terms, abstract concepts, and the necessity for construction and reconstruction of mental models (Newcombe et al., 2009). Fortunately, research has shown that learning can improve through the deployment of key cognitive and metacognitive processes such as planning, monitoring, and through the use of appropriate learning strategies (Azevedo, 2009; Dunosky & Metcalfe, 2009; Hacker, Dunlosky, & Graesser, 2009; Pintrich, 2000; Winne, 2011; Winne & Hadwin, 2008; Zimmerman & Schunk, 2011). These processes, also called self-regulated learning (SRL) processes, are based on the assumption that learners actively monitor and control their learning to aid in deeper processing of the material (Azevedo & Witherspoon, 2009).

Self-regulated learning is an active and constructive process that involves learners' ability to build on their understanding of a topic by using planning, monitoring, and learning strategies, and by regulating key aspects of cognition, behavior, motivation, and affect in order to achieve some desired learning goal (Azevedo & Witherspoon, 2009; Boekaerts, Pintrich, & Zeidner, 2000; Koriat, Ma'ayan, & Nussinson, 2006; Pintrich, 2000; Zimmerman & Schunk, 2011). More specifically, learning of complex science topics necessitates learners to effectively self-regulate their learning by *metacognitively monitoring* their emerging understanding of a given topic (Burkett & Azevedo, 2012; Graesser et al., 2007; Shapiro, 2008). Most research on the topic of metacognitive monitoring focuses primarily on *metacognitive judgments* (see Dunosky and Metcalfe (2009) for a recent review), which occur before, during, and after learning has taken place, as learners continually assess their emerging understanding of the material.

---

[*] Corresponding author. Address: Department of Psychology, Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152, United States.
E-mail address: dchuncey@memphus.edu (A.C. Strain).

0361-476X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cedpsych.2012.08.001

There are three metacognitive judgments that are most commonly examined in SRL research. These include ease of learning (EOL), judgments of learning (JOL), and retrospective confidence judgments (RCJs) (Dunosky & Metcalfe, 2009; Leonesio & Nelson, 1990; Nelson & Narens, 1990). Ease of learning judgments occur *before* learning and involve preemptively determining how easily a given topic can be learned. They occur in the prospective phase of learning and are assumed to help learners establish goals, subgoals, and allocation of study-time, and can be used as a baseline comparison for future metacognitive judgments. Judgments of learning occur *during* learning when learners attempt to assess their emerging understanding of the topic, and are predictive of subsequent learning performance (Jang & Nelson, 2005). Retrospective confidence judgments occur *after* learning has taken place when learners predict how likely it is that their responses to evaluative items were correct.

Examining the use of metacognitive monitoring processes can provide several insights into how learners regulate their learning. However, an equally important component that is gaining attention in the domain of SRL is the role of affect (Brosch, Pourtios, & Sander, 2010; Frijda, 2009; Izard, 2007; Schwarz, 2011; Stein, Hernandez, & Trabasso, 2008). There are many terms that are used to describe learners' affective experiences, such as basic emotions (Ekman, 1992), moods (Bless, 2000; Bower & Forgas; 2000; Isen, 2001, 2010; Schwarz & Clore, 1983), affective states (D'Mello & Graesser, 2011a, 2012) and academic emotions (Pekrun, 2010). Within the category of academic emotions, there are various other terms such as acheivement emotions, topic emotions, social emotions, and epistimic emotions (Pekrun, 2010). Although each of these terms are distinct and important in their own way, this article uses the term *affect* or *affective states* broadly to encapsulate the feelings and emotions that arise during brief learning episodes (30 min to 2-h). This consists of reactions to specific learning events that vary in intensity but are relatively brief, lasting for a few seconds to a few minutes (D'Mello & Graesser, 2011a; Rosenberg, 1998). What is not meant by *affect*, however, are moods (longer term affective experiences that are not directed at any particular event), affective traits (predispositions in affective responding), or motivational states. Previously published papers offer justification for this conceptualizaton of affect (Baker, D'Mello, Rodrigo, & Graesser, 2010; Calvo & D'Mello, 2011; Conati & Maclaren, 2009; Rosenberg, 1998; Woolf et al., 2009).

Part of the challenge of learning about conceptually-rich domains such as science, technology, engineering, and mathematics is that these domains are rife with affect-eliciting factors such as complexity of the learning materials, uncertainty about how to proceed when faced with obstacles to learning, and the fear of performing poorly on subsequent evaluations. These negative factors can interfere with learners' ability to effectively regulate their learning. Although many conceptual models of SRL focus on learners' use of metacognitive monitoring and control processes to regulate their learning (Azevedo, Moos, Johnson, & Chauncey, 2010; Dunlosky & Theide, 2004; Dunosky & Metcalfe, 2009; Metcalfe, 2002; Zimmerman & Schunk, 2011), the role of affect during learning has, until recently, received somewhat less attention. Existing models that address learners' affect tend to focus primarily on how affect broadly impacts motivational, metacognitive, and cognitive processes. For example, increases in self-satisfaction (a positively valenced affective state) are correlated with enhanced motivation and effort, while decreases are associated with diminished effort (Schunk, 2001). Self-efficacy is also associated with the use of varied study methods in order to discover new avenues for self-improvement (Zimmerman, 2002), and is related to learners' use of SRL strategies (Braten, Samuelstuen, & Stromso, 2004). Other models of SRL explore the role of affective processes on motivation (Boekaerts, 2009; Pintrich, 2000), goal orientation (Harachiewicz, Barron, Pintrich, Elliot, & Thrash, 2002), interest (Pintrich & Schunk, 2002; Wigfield, Eccles, Schiefele, Roesner, & Davis-Kean, 2006), and the relationship between products (i.e., learning outcomes) and standards (i.e., learners' evaluations of optimal end states) (Winne & Hadwin, 2008).

While these models focus primarily on broad effects of affect on a number of outcome variables, the present research diverges from, but builds upon, these models by attempting to uncover the intricate relationship between affect, SRL (specifically metacognitive components of SRL), and learning outcomes. Investigation into the relationship among these processes is essential, because there is a complex interplay between cognitive and affective processes during learning and problem solving (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello & Graesser, 2011b; Daniels, Stupnisky, et al., 2009; Daniels, Pekrun, et al., 2009; Linnenbrink, 2006; Meyer & Turner, 2006; Pekrun, 2010; Schutz & Pekrun, 2007; Zeidner, 2007). Affect operates throughout cognitive processes such as causal reasoning, deliberation, goal appraisal, and planning. Flexibility, creative thinking, efficient decision-making, and conceptually-driven relational thinking have been linked to positive affect, while negative affect has been associated with localized attention and stimulus-driven processing (Clore & Huntsinger, 2007; Fielder, 2001; Fredrickson & Branigan, 2005; Isen, 2008; Schwarz, 2011). Affect can also have a serious impact on learners' comprehension and performance on evaluative measures (Zeidner, 2007).

Importantly, it is perhaps not the affective states themselves, but the cognitive and metacognitive activities that accompany their experience that are predictive of learning. This leads to the critical question of how affect influences these metacognitive and cognitive processes, a question that motivated the present research.

2. Theoretical framework, hypotheses, and present research

The current research adopts an appraisal theoretic framework to describe the antecedents of learners' affective states. Contemporary theories of affect posit that cognitive appraisals of physiological changes are one prominent way that affective states arise (Barrett, Mesquita, Ochsner, & Gross, 2007; Lazarus, 1991; Mandler, 1975, 1999; Ortony, Clore, & Collins, 1988; Russell, 2003; Schachter & Singer, 1962; Stein & Levine, 1991). The specific affective states that arise depend on an individual's unconscious or conscious appraisals (i.e., evaluations) of the situation that presumably caused the physiological change along dimensions such as novelty, goal-alignment, agency, coping potential, and availability of a plan (Cacioppo, Klein, Berntson, & Hatfield, 1993; Duffy, 1962; Karsdorp, Kindt, Rietveld, Everaerd, & Mulder, 2009; Ortony et al., 1988; Schachter & Singer, 1962; Valins, 1966).

Building on this research foundation, the fundamental question addressed in this article is how affect influences learners' metacognitive judgments and performance. We conducted an experiment that used a false-biofeedback methodology (Kirsch & Lynn, 1999; Schachter & Singer, 1962; Valins, 1966) to induce physiological arousal (and resultant affective states) by presenting learners with audio stimuli of false heart beats that were either baseline, like those that would be experienced if an individual was in a neutral state, and accelerated, like those that would be experienced in a moment of excitement or fear. In some trials we presented learners with no auditory stimulus; these trials served as the control trials.

One key concept related to false biofeedback is the occurrence of physiological alignment with the presented auditory stimulus. Previous research has demonstrated that participants' physiological responses will align with false heart rate biofeedback and false skin conductance biofeedback (Ehlers, Margraf, Roth, Taylor, & Birbaumer, 1988; Holroyd et al., 1984; Lichstein & Hoelscher, 1989). That is, when participants hear an accelerated heart rate, their own heart rate increases, and when participants see a graph depicting an increase in their skin conductance over time, their own skin conductance will increase. These responses, in turn, are associated with the appraisal that one is in a state of elevated physiological arousal, and a search for the cause of that arousal is initiated.

In the context of learning, there are two possible results of this search for the cause of physiological arousal. It is possible that such a search would direct attentional resources away from the learning task and onto the cause of the arousal, which might impair learning. On the other hand, it is also possible that engaging in such a search would cause a learner to more closely examine the events leading up to the feeling of arousal, which might result in higher metacognitive awareness.

As such, our goal of using false biofeedback was to examine how appraisals of physiological feedback influence affective states, metacognitive judgments, and performance (specifically, learning outcomes). That is, we sought to determine whether the appraisal of physiological arousal can influence shifts in learners' self-reported affect and the confidence with which metacognitive judgments are made, and facilitate or impair learning. Because affective states arise spontaneously and decay at varying intervals (D'Mello & Graesser, 2011a), evaluating their relationship with such processes can be murky and difficult to unveil. Therefore, although we acknowledge that a potential limitation to the false-biofeedback methodology is its lack of ecological validity, its usefulness in the current research is its ability to uncover precise, experimentally controlled relationships among affect, metacognition, and performance. Using appraisal theories of emotion as a guide, we developed a number of hypotheses about the influence of false biofeedback on self-reported affect, metacognitive judgments, and performance.

2.1. Alarm vs. alert hypothesis

The *alarm hypothesis* predicted that false biofeedback would be unpleasant or distracting, and thus would be associated with feelings of high arousal and unpleasant affect, which would lead to less confident metacognitive judgments and decreased performance (Eysenck, Derekshan, Santos, & Calvo, 2007). The *alert hypothesis* predicted that false biofeedback would be associated with feelings of alertness, engagement, or interest, which would lead to more confident metacognitive judgments and increased performance.

The *alarm hypothesis* is based on empirical findings that suggest that physiological arousal will be appraised as being indicative of something problematic, resulting in the experience of negative affective states (Schwarz, 2000, 2011; Schwarz & Clore, 2003; Schwarz & Skurnik, 2003). This negative affect can lead learners to focus on negative self- and task-related information (Elliot & Thrash, 2002; Linnenbrink & Pintrich, 2002; Linninbrink, 2007), avoidance achievement goals (Pekrun, Elliot, & Maier, 2009; Pekrun, Frenzel, Goetz, & Perry, 2007), and decreased performance (Church, Elliot, & Gable, 2001; Eysenck et al., 2007; Zeidner, 2007). Negative affect can also impact metacognitive judgments (Efklides & Petkaki, 2005), such as making learners less confident in their current understanding of a topic or in their performance following an evaluation. Hence, according to the *alarm hypothesis*, participants should appraise false biofeedback as an indication that they are in a negative affective state, which would lead them to report *high* physiological arousal and *negative* valence, make *less confident* metacognitive judgments, and achieve *lower performance*, compared to when they received no biofeedback.

Although an extensive body of research on test-anxiety has demonstrated that learners often equate elevated arousal with stress or anxiety (Pekrun, Goetz, Titz, & Perry, 2002; Schutz & Pekrun, 2007; Zeidner, 2007), an alternate position is that physiological arousal can elicit positive feelings such as interest or engagement. This possibility was the foundation for the *alert hypothesis*. According to this hypothesis, participants would make the appraisal that arousal was an indication that they were alert, engaged, and interested in the learning task. These states, in the context of learning, are indicative of positive activating affective states (Pekrun, 2010). Considerable empirical research has demonstrated that positive affect can facilitate decision making (Fredrickson & Branigan, 2005; Isen, 2010), enhance problem solving (Clore, 1992; Fredrickson, 2001; Isen, 2004), encourage the adoption of performance-approach and mastery goals (Pekrun et al., 2002, 2009), and improve performance (Zeidner, 2007). As it relates to metacognition, positive affect can cue learners that their learning is under control, meaning they might feel that the material is comprehensible, that a sufficient understanding of the topic has been attained, and that they are capable of performing well on a subsequent evaluation (Efklides, 2006; Schwarz & Clore, 2003). Therefore, the *alert hypothesis* predicted that when participants received false biofeedback, they would self-report *high* physiological arousal and *positive* valence, make *more confident* metacognitive judgments, and achieve *higher* learning performance, compared to when they received no feedback.

2.2. Differential biofeedback hypothesis

In addition to the *alarm and alert hypotheses,* it is also possible that different levels of false biofeedback might lead to different trajectories of affect, metacognition, and performance. It is possible that (in line with the alarm hypothesis) *accelerated* biofeedback might lead to appraisals of distress or anxiety, or could distract attentional resources away from the learning task and onto participants' physiological arousal and affective states, which would result in negative metacognitive judgments and poor performance. In contrast, (and in line with the alert hypothesis) *baseline* biofeedback might cue participants that the learning activity is going well and that physiological arousal is low (i.e., *My heart is not racing, which means I am calm, so I must be understanding this material*). These appraisals, in turn, might lead to positive metacognitive judgments and high performance. As such, the *differential biofeedback hypothesis* predicts that affective states, metacognitive judgments, and learning performance will differ significantly when participants are presented with accelerated vs. baseline biofeedback.

2.3. Question depth hypothesis

Aspects of the task are also expected to influence how perceived physiological arousal impacts affect, metacognition, and performance. In particular, there is a question of whether learners react differently to false biofeedback when they read a text and have to answer a difficult question about that text, rather than an easier question. To examine this issue, we presented learners with *text based* questions, which were conceptually simple and required shallow understanding of the topic, and *inference* questions, which required a deep conceptual understanding of the topic in order to be answered correctly (Graesser, Ozuru, & Sullins, 2010; Graesser & Person, 1994). Our prediction, called the *question depth hypothesis,* was that we would observe the strongest effects of false biofeedback when learners were asked to answer inference questions, since these questions contain within them potentially affectively-charged factors such as difficult phrasing, long or unfamiliar words, and varying levels of abstractness. Any effect of false biofeedback would be diminished when learners were asked to answer a text based question, since these questions could be answered by simply locating a key sentence within each text and, therefore, are inherently less anxiety-provoking.

2.4. Self vs. recording hypothesis

One final question pertains to whether learners must actually believe that false biofeedback is indicative of their *own* physiological arousal in order to experience positive or negative affect. That is, must they believe that they are hearing their own heart beat, rather than a false heart beat *intended* to be perceived as their own? Although a considerable amount of research (Ehlers et al., 1988; Holroyd et al., 1984; Lichstein & Hoelscher, 1989) has indicated that actual physiological arousal will align with intended arousal from the false biofeedback manipulations, an alternative view is that alignment might fail to occur. Learners might detect a mismatch between their own physiological arousal and the false biofeedback being presented, which may cause them to believe that the false biofeedback was, in fact, false (i.e., *this has* got *to be a recording*). If this is the case, how might learners' affective states, metacognitive judgments, and learning performance be affected (if they are affected at all)? To test this question, we conducted a second experiment in which participants in one condition were explicitly informed that the biofeedback was a recording and was in no way an indication of their own physiological arousal. Participants in the second condition, similar to participants in Experiment 1, were instructed that the biofeedback *was* indicative of their physiological arousal. The hypothesis, which we call the *self* vs. *recording hypothesis*, was that there would be significant differences between learners who believed that the false biofeedback to their own heart beat and learners who knew that the false biofeedback was fake.

3. Experiment 1

3.1. Method

3.1.1. Participants

Fifty undergraduate students from a large urban university in the US participated in this experiment. The participants' mean age was 23.3 years (*SD* = 7.13), and there were 34 females (68%) in the sample. There were 54% Caucasians, 44% African Americans, and 2% were Hispanic. All participants received $20 for participating in the experiment.

3.1.2. Design

The experiment used a 3 (Biofeedback: Accelerated, Baseline, and No biofeedback) × 2 (Question Type: Text based and Inference) within-subjects design.

3.1.3. Stimuli and software

A linearly-structured self-paced instructional system described in D'Mello, Lehman, and Person (2011) presented multimedia content comprised of 24 slides about the human circulatory system (see Fig. 1). The interface was composed of a large window for presenting content (on the left), and a small window where participants could respond to prompts and answer questions (on the right).

Participants received accelerated false biofeedback for eight randomly selected slides, baseline false biofeedback for another eight randomly selected slides, and no false biofeedback for the remaining eight slides. The type of false biofeedback (accelerated, baseline, or control) was randomly presented across trials. Our rationale for presenting false biofeedback randomly (rather than in blocks) stemmed from the concern that hearing one type of false biofeedback for a block of eight consecutive trials might seem unnatural and cue participants that the biofeedback was not indicative of their own arousal. The random presentation of false biofeedback was assumed to be more reflective of the natural shifts in physiological processes that often occur throughout a lengthy learning episode.

The two auditory stimuli used in this experiment were presented binaurally through headphones. During accelerated trials, participants heard a digital recording of an accelerated human heart beat (approximately 100 BPM), and during baseline trials, they heard a digital recording of a resting human heart beat (approximately 70 BPM). During control trials, no auditory stimulus was presented. These stimuli were initiated when participants opened a content slide and the audio streamed continuously until they navigated away from the slide by clicking a navigational arrow at the bottom of the screen.

3.1.4. Apparatus

A Reebok™ Fit Watch 10 s strapless heart rate monitor was worn around participants' non-dominant wrist. This heart rate monitor was designed to detect and display the wearer's heart rate. However, because previously-recorded accelerated and baseline heart beats were presented to participants (rather than their own heart beat), this function was not used for this experiment. The purpose of the watch was simply to cause participants to believe that their heart beat was being recorded and presented back to them during the session.

3.1.5. Paper and pencil materials

The Affect Grid (Russell, Weiss, & Mendelsohn, 1989) is a validated measure of affect with adequate reliability (Cronbach's α score of .85), convergent validity (correlations of .90 or higher with similar scales of affect), and discriminant validity (correlations of .20 or less with dissimilar scales of affect). It is a single item affect measurement instrument consisting of a 9 × 9 (valence × arousal) grid; these are the primary dimensions that underlie affective experiences (Barrett, 2009). The arousal dimension ranges from low arousal/sleepy (1) to high arousal/active (9), while the valence dimension ranges from unpleasant feelings (1) to pleasant feelings (9). Participants indicated their current affective state by marking an X at the appropriate location on the grid (see Fig. 2).

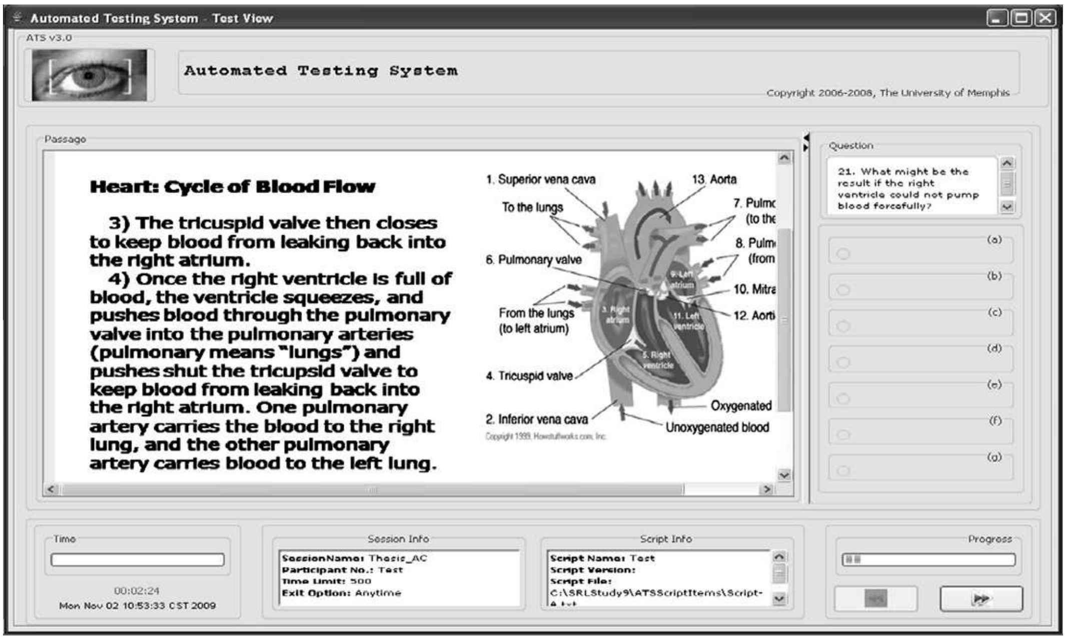
3.1.6. Content covered in the session

The multimedia content presented in the learning session consisted of 24 slides about the human circulatory system. Each slide contained a passage of text and a corresponding diagram. The passages were similar in length, with an average of 82.3 (*SD* = 19.7) words per slide. They had a Flesch-Kincaid score of 9.0. The corresponding diagrams were presented adjacent to the text and provided illustrative examples of the content being presented on each slide.

Either a text based (i.e., *What is a primary role of the circulatory system?)* or inference question (i.e., *With age, the aortic valve sometimes accumulates deposits of calcium, the valve becomes stiffened, and the opening narrows. What might be an effect of this situation?)* was presented along with each slide, and participants were provided with four multiple choice foils for answering each question. These four foils consisted of the target (the correct response to the question), a near-miss (an option that sounded correct but was not), a thematic miss (an option that followed the theme of the content but was not actually related to the question) and a miss (an option that was not at all related). Participants answered each question by selecting one of these four options. Text based questions had an acceptable reliability, with a Chronbach's alpha of 0.72. Inference questions had a somewhat lower, but acceptable, reliability (α = .64).

3.1.7. Manipulation check

A manipulation check was developed to determine whether participants believed that the false biofeedback was indicative of



Heart: Cycle of Blood Flow

3) The tricuspid valve then closes to keep blood from leaking back into the right atrium.

4) Once the right ventricle is full of blood, the ventricle squeezes, and pushes blood through the pulmonary valve into the pulmonary arteries (pulmonary means "lungs") and pushes shut the tricupsid valve to keep blood from leaking back into the right atrium. One pulmonary artery carries the blood to the right lung, and the other pulmonary artery carries blood to the left lung.

21. What might be the result if the right ventricle could not pump blood forcefully?

00:02:24
Mon Nov 02 10:53:33 CST 2009

SessionName: Thesis_AC
Participant No.: Test
Time Limit: 500
Exit Option: Anytime

Script Name: Test
Script Version:
Script File:
C:\SRLStudy9\ATSScriptItems\Script-

Fig. 1. Screen shot of the learning environment.

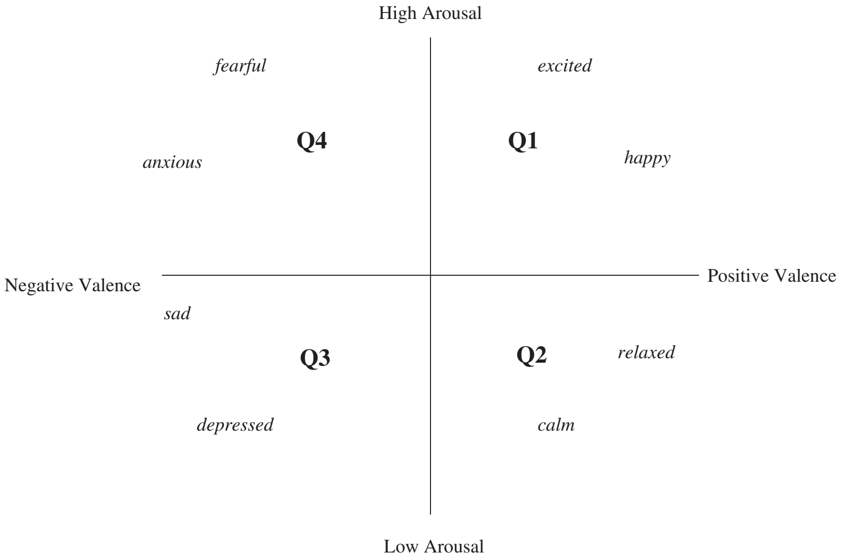


Fig. 2. The Affect Grid (Russell et al., 1989) divided into four quadrants. The *X* axis represents the dimension of valence, ranging from negative to positive. The *Y*-axis represents the dimension of arousal, ranging from low to high. Positive affect states are in Q1 and Q2, and negative states are in Q3 and Q4. Words in italics indicate possible locations of specific affective states.

their own physiological arousal, or whether they were conscious of the fact that the false biofeedback was indeed false and, thus, was not indicative of their physiological arousal. The assumption of this manipulation check was that if participants reported believing that the false biofeedback was their own, then we could reasonably infer that physiological alignment did in fact occur, and thus, that our manipulation had the intended effect. At the conclusion of the experiment, participants engaged in a verbal interview with the researcher. There was a concern that directly asking participants whether they believed the heart beats they heard were their own might bias their response. Thus, participants were told, "*Sometimes the audio gets messed up on these computers and causes the heart beats participants hear during the experiment to be a little delayed. I have another participant coming in after you and will need to calibrate the system if it's not working properly. Did you think that the heart beats you heard during the experiment were generally correct? I mean, for example, if you heard a fast heart beat did you think your heart really was beating fast?*" Participants responded verbally to this question, and their responses were recorded by the experimenter.

3.1.8. Procedure

Participants were tested individually in a 1.5 h session. After completing some demographic measures, participants read the prescribed instructions for using the Affect Grid (see Russell et al., 1989). Next, participants were instructed to strap the wristwatch around their non-dominant wrist before beginning the learning session. Participants were instructed that throughout the learning session their heart rate would be collected via the wristwatch and transmitted wirelessly to a sensor mounted on the desk beside them. The experimenter explained that a program installed on the computer would collect, record, and randomly present their heart rate through headphones during some trials. Before beginning the session, the experimenter stated, "The program is designed to only present your heart rate to you while you are reading a content slide. This means that you will begin hearing your heart rate when you navigate to a content slide, and will cease to hear it when you navigate away. Remember that you will not hear your heart rate in every trial, so do not fret if you open a content slide and do not hear anything. That is completely normal. Everyone's heart rate is different, and yours might fluctuate throughout the experiment. If at any time hearing your heart rate makes you anxious or uncomfortable, please alert the experimenter."

The learning session took place over 24 trials, with each trial consisting of multiple steps. First, participants viewed either a text based or inference question related to the content. At this time participants were not shown the four multiple choice options for answering the question. After reading the question, participants were asked to make an ease of learning (EOL) judgment. That is, they were asked to indicate how easily they could learn the material required to answer the question they had just seen. Participants made this judgment on a six-item scale ranging from 1 (*I strongly feel this will be difficult to learn*) to 6 (*I strongly feel this will be easy to learn*).

Next, participants had as much time as desired to read the content and study the corresponding diagram on the content slide. Upon opening the content slide, the learning environment presented either accelerated, baseline, or no biofeedback through participants' headphones. This false biofeedback played continuously until participants navigated away from the slide, and terminated as soon as the next slide appeared.

When participants navigated to the next slide, they were prompted to indicate how well they understood what they had just read by making a judgment of learning (JOL) on a six item scale ranging from 1 (*I strongly feel I do not understand*) to 6 (*I strongly feel I understand*). Following the JOL prompt, the text based or inference question was presented again and participants were prompted to answer the question by selecting one of the four multiple choice foils. Next, participants were prompted to rate the accuracy of their answer by making a retrospective confidence judgment (RCJ) on a six-item scale ranging from 1 (*I strongly feel my answer was incorrect*) to 6 (*I strongly feel my answer was correct*). For the final step in each trial, participants were prompted to self-report their current level of valence and arousal on the Affect Grid. The completion of the Affect Grid marked the end of one trial. This multi-step process occurred for all 24 trials within the self-paced learning session.

It should be noted that there was a concern that presenting false-biofeedback continuously throughout each trial would cause carry-over effects from trial to trial. As such, we felt that it was important to only present false biofeedback while participants were reading the learning materials. Because several tasks were completed between the termination of false biofeedback in one trial and the beginning of the next trial (i.e., JOL, answering the question, RCJ, Affect Grid), we felt that the carry-over effect was unlikely. To confirm this assumption, we correlated arousal and valence on trial N with arousal and valence on trial $N+1$ when there was a change in biofeedback between trials and found no significant correlations, suggesting that carry-over was not a major concern.

A heart rate recognition task was administered upon completion of the experiment. This task was intended ensure that participants were able to differentiate between accelerated and baseline false biofeedback. Participants were randomly presented with 5-s samples of accelerated and baseline false biofeedback (five accelerated and five baseline). They were instructed to listen to each sample and determine which type of biofeedback they just heard by clicking a button labeled either *accelerated* or *baseline*. All 50 participants correctly identified all 10 samples. Lastly, all participants engaged in the verbal interview with the researcher to ensure that the experimental manipulation was successful, and were subsequently debriefed.

3.2. Results

Our analyses examined the effect of false biofeedback and question type on learners' *affective states, metacognitive judgments*, and *performance*. Participants' scores along the arousal and valence dimensions on the Affect Grid were used as a measure of their affective states. The metacognitive judgments examined were participants' ease of learning judgments, judgments of learning, and retrospective confidence judgments. We examined participants' accuracy on the multiple-choice questions as a measure of their performance. Descriptive statistics for these variables can be found in Table 1. A Bonferroni correction was applied to all analyses, and an alpha level of 0.05 was adopted for significance testing for all subsequent analyses unless specified otherwise.

3.2.1. Manipulation check

Of the 50 participants in the study, 46 participants (92%) reported that they believed the heart beats they heard were correct, indicating that they believed the false biofeedback was indicative of their own physiological arousal. Believing that the false biofeedback was a reflection of one's own heart rate was the essential component of the experimental manipulation. As such, participants who reported being skeptical of the false biofeedback were excluded from analysis, yielding a sample size of 46 participants.

3.2.2. Affective states

Affect theories posit that arousal and valence are interrelated in a highly complex and inextricable fashion (Linninbrink, 2007; Russell, 1980, 2003). As such, we conducted a 3 × 2 repeated measures multivariate analysis of variance (MANOVA) to first evaluate the overall effect of false biofeedback and question type on valence and arousal together. For each of the following analyses, *false-biofeedback* will be referred to as *biofeedback*, and the type of question participants were asked to answer (text based or inference) will be referred to as *QT*.

The multivariate model revealed a significant main effect of biofeedback, Wilks' $\lambda = 0.853$, $F(4,178) = 4.03$, partial $\eta^2 = 0.077$, on valence and arousal jointly. There was also a significant effect of QT, Wilks' $\lambda = 0.532$, $F(2,44) = 21.15$, partial $\eta^2 = 0.486$. There was no significant biofeedback × QT interaction. Because the overall model was significant, subsequent univariate analyses of variance (ANOVAs) were examined to evaluate the effect of biofeedback and question type on arousal and valence separately.

3.2.2.1. Arousal. We found a significant main effect for biofeedback on participants' reported arousal, $F(2,90) = 7.65$, $MSE = 0.443$, partial $\eta^2 = 0.135$. Bonferroni post hoc tests indicated that participants reported significantly more arousal when they heard accelerated ($M = 5.28$, $SD = 1.32$, $d = 0.30$), or baseline biofeedback ($M = 5.34$, $SD = 1.47$, $d = 0.32$), than when they heard no biofeedback ($M = 4.88$, $SD = 1.38$) (Accelerated = Baseline > Control).We also found a significant main effect for QT, $F(1,45) = 6.92$, $MSE = 0.704$, partial $\eta^2 = 0.124$, with participants reporting significantly more arousal when they were asked to answer an inference question ($M = 5.27$, $SD = 1.48$) than a text based question ($M = 5.02$, $SD = 1.27$), $d = 0.18$. There was no significant biofeedback × QT interaction.

Table 1
Descriptive statistics for dependent variables (Experiment 1).

	Text-based question			Inference question		
	Accelerated *M* (*SD*)	Baseline *M* (*SD*)	Control *M* (*SD*)	Accelerated *M* (*SD*)	Baseline *M* (*SD*)	Control *M* (*SD*)
Affective						
Arousal	5.36 (1.61)	5.34 (1.38)	5.11 (1.75)	5.10 (1.22)	5.20 (1.63)	4.76 (1.23)
Valence	6.41 (1.18)	6.38 (1.19)	6.48 (1.07)	5.78 (1.37)	6.06 (1.19)	5.57 (1.12)
Metacognitive						
EOL	4.71 (1.01)	4.72 (.891)	5.20 (.767)	4.39 (.912)	4.12 (1.08)	4.08 (1.01)
JOL	5.11 (.874)	5.28 (.679)	5.39 (.661)	4.70 (.871)	4.59 (.812)	4.28 (.939)
RCJ	5.41 (.551)	5.32 (.566)	5.35 (.789)	4.82 (.741)	4.50 (.632)	4.08 (.810)
Cognitive						
Performance	.847 (.213)	.841 (.221)	.852 (.243)	.837 (.211)	.736 (.199)	.554 (.192)

Note: Performance scores indicate the proportion of correct answers out of all possible answers.

3.2.2.2. Valence. We failed to find a significant main effect for biofeedback, or a significant biofeedback × QT interaction, indicating that biofeedback had no significant impact on participants' experience of positive or negative states. We did find a significant main effect for QT, $F(1,45) = 41.5$, $MSE = 0.693$, partial $\eta^2 = 0.458$, which indicated that participants experienced more positive valence when they were asked to answer a text based question ($M = 6.40$, $SD = 0.982$) rather than a difficult inference question ($M = 5.81$, $SD = 0.994$), $d = 0.60$.

3.2.2.3. Mapping valence and arousal on the Affect Grid. We mapped the combination of participants' self-reported valence and arousal onto the Affect Grid to explore how they *jointly* differed across accelerated, baseline, and no biofeedback. The Affect Grid, which is derived from Russell's (1980) Circumplex model of affect, is divided into four quadrants, each representing a different combination of valence and arousal (see Fig. 2). For example, the top-right quadrant represents High Arousal/Positive Valence and a prototypical affective state that might be classified in this quadrant is excitement or joy. The bottom-right quadrant, on the other hand, represents Low Arousal/Positive Valence. An affective state that might be represented in this quadrant is relaxation or calmness.

In our sample, as Fig. 3 demonstrates, participants' valence and arousal mapped onto the High Arousal/Positive Valence quadrant (Q1) when they received accelerated and baseline biofeedback. When they received no biofeedback, their self-reported valence remained positive (above neutral), but their arousal dropped below neutral, which placed their combined valence and arousal in the Low Arousal/Positive Valence quadrant (Q2).

In order to statistically test whether false biofeedback caused a change in participants' affect (i.e., change from the neutral state), we computed the Euclidian distance between self-reported affect (a point in the valence-arousal space) from the neutral point (coordinates [5,5] on the Affect Grid) during the accelerated, baseline, and no biofeedback trials. A one-factor repeated measures ANOVA yielded a significant main effect for biofeedback, $F(2,44)$, $MSE = 0.255$, partial $\eta^2 = 0.116$. A Bonferroni post hoc test revealed that participants' Euclidean distance from neutral was significantly greater for accelerated ($M = 1.84$, $SD = 0.968$, $d = 0.33$) and baseline trials ($M = 1.82$, $SD = 0.981$, $d = 0.31$) than control trials ($M = 1.52$, $SD = 0.944$). There were no differences in Euclidean distance between accelerated and baseline trials. This finding demonstrates that although we failed to find a significant main effect for false biofeedback on valence, our manipulation did evoke significant shifts from the neutral state. Specifically, during false biofeedback trials, participants' affective states mapped onto the positive activating quadrant of the Affect Grid, while during control trials they mapped onto the positive deactivating quadrant.

3.2.3. Metacognitive judgments

We conducted separate repeated measures ANOVAs for each of the three metacognitive judgments reported by participants across all 24 trials within the learning session. Because EOL judgments occurred prior to the presentation of false biofeedback, we conducted a one-way repeated measures ANOVA to compare only the effect of QT on participants' EOL judgments. We conducted a 3 × 2 (biofeedback × QT) repeated measures ANOVA to compare the effect of biofeedback and QT on participants' JOLs and RCJs. Descriptive data for each of these variables is presented in Table 1.

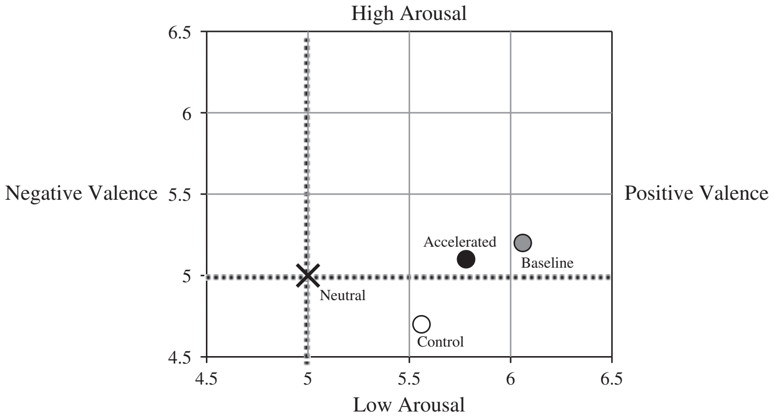


Fig. 3. Mean combined valence and arousal scores across biofeedback for inference questions (Experiment 1).

3.2.3.1. Ease of learning judgments. The analysis of participants' EOL judgments indicated that there was a significant main effect for QT, *F*(1,45) = 75.1, *MSE* = 0.178, partial η^2 = 0.605. Bonferroni post hoc analyses revealed that participants made significantly higher EOL judgments for text based questions (*M* = 4.80, *SD* = 0.792) than inference questions (*M* = 4.10, *SD* = 0.943), *d* = 0.81. That is, participants believed that content associated with text based questions would be significantly easier to learn than content associated with inference questions.

3.2.3.2. Judgments of learning. We failed to find a significant main effect of biofeedback on participants' JOLs. However, we did find a significant effect for QT, *F*(1,45) = 145.1, *MSE* = 0.278, partial η^2 = 0.748, revealing that participants made significantly higher JOLs when they answered text based questions (*M* = 5.30, *SD* = 0.662) than inference questions (*M* = 4.50, *SD* = 0.775), *d* = 1.12.

We also found a significant biofeedback × QT interaction, *F*(2,44) = 16.39, *MSE* = 0.189, partial η^2 = 0.25 (see Fig. 4). Bonferroni post hoc analyses revealed that for text based questions, participants made significantly higher JOLs when they heard no biofeedback or baseline biofeedback compared to when they heard accelerated biofeedback, (Control = Baseline > Accelerated). For inference questions we found the opposite pattern. That is, participants made significantly higher JOLs when they heard accelerated or baseline biofeedback than when they heard no biofeedback (Baseline = Accelerated > Control). Therefore, it seems that the effect of false biofeedback on JOLs was influenced by question type.

3.2.3.3. Retrospective confidence judgments. For participants' RCJs, we found a significant main effect for biofeedback, *F*(2,44) = 11.5, *MSE* = 0.342, partial η^2 = 0.190. Bonferroni post hoc tests indicated that participants were significantly more confident in the accuracy of their answers when they heard accelerated biofeedback (*M* = 5.12, *SD* = 0.584) than when they heard baseline biofeedback (*M* = 4.91, *SD* = 0.512), *d* = 0.38, which, in turn, was greater than when they received no biofeedback (*M* = 4.72, *SD* = 0.647), *d* = 0.67 (Accelerated > Baseline > Control). The effect of QT was also significant, *F*(1,45) = 217.2, *MSE* = 0.275, partial η^2 = 0.81, with participants reporting significantly higher RCJs when they answered a text based question (*M* = 5.36, *SD* = 0.454) than an inference question (*M* = 4.47, *SD* = 0.587), *d* = 1.69.

We also found a significant biofeedback × QT interaction, *F*(2,44) = 10.9, *MSE* = 0.331, partial η^2 = 0.183 (see Fig. 5). Bonferroni post hoc tests revealed that when participants were asked to answer a text based question, their RCJs did not differ significantly across accelerated, baseline, or no false biofeedback (Accelerated = Baseline = Control). The interesting pattern (Accelerated > Baseline > Control) only emerged when participants were asked to answer a more conceptually challenging inference question.

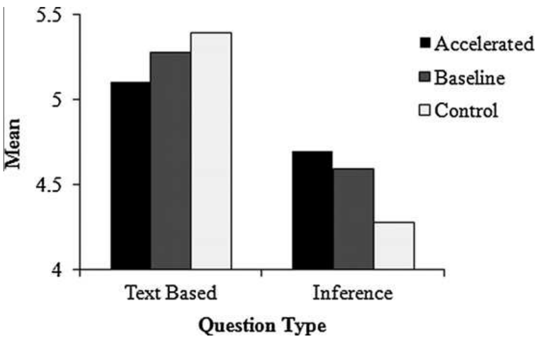


Fig. 4. Means of participants' JOLs by biofeedback and question type (Experiment 1).

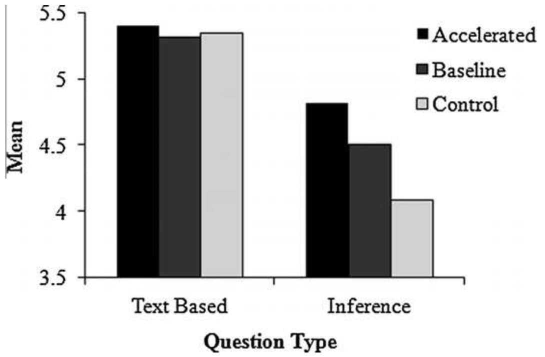


Fig. 5. Means of participants' RCJs by biofeedback and question type (Experiment 1).

3.2.4. Performance

A 3 × 2 repeated measures ANOVA was conducted to compare the effect of biofeedback and QT on participants' performance. We found a significant main effect of biofeedback, *F*(2,44) = 14.2, *MSE* = 0.032, partial η^2 = 0.225. Post hoc tests revealed an Accelerated = Baseline > Control pattern, indicating that participants performed significantly better when they heard accelerated (*M* = 0.83, *SD* = 0.19, *d* = 0.80) or baseline biofeedback (*M* = 0.79, *SD* = 0.14, *d* = 0.62) than when they heard no biofeedback at all (*M* = 0.70, *SD* = 0.15). We also found a significant effect for QT, *F*(1,45) = 28.9, *MSE* = 0.047, partial η^2 = 0.371, with participants performing significantly better on text based questions (*M* = 0.84, *SD* = 0.15) than inference questions (*M* = 0.71, *SD* = 0.16), *d* = 0.84.

There was a significant biofeedback × QT interaction as well, *F*(2,44) = 15.7, *MSE* = 0.33, partial η^2=.243(see Fig. 6). Similar to the findings from participants' RCJs, this interaction revealed that when participants were asked to answer a text based question their learning performance did not differ significantly across the three types of biofeedback (Accelerated = Baseline = Control). However, when they were asked to answer an inference question, participants' learning performance did significantly differ, with scores significantly higher when participants heard accelerated or baseline biofeedback compared to no biofeedback (Accelerated = Baseline > Control).

3.2.5. Relationships among key dependent variables

Thus far, we have examined the effect of false biofeedback and question type on learners' affect, metacognition, and performance. However, a broader goal of this research also involved understanding how these complex processes interact during learning. We addressed this goal by conducting canonical correlation analyses using 10 metacognitive and affective variables (Arousal, Valence,

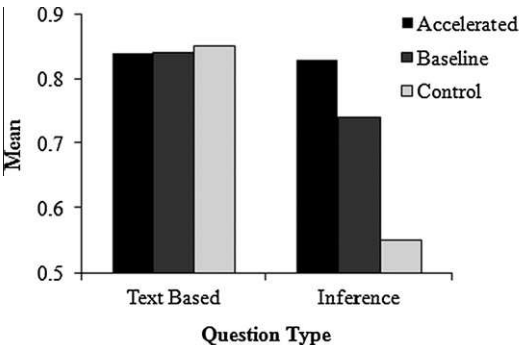


Fig. 6. Means of participants' performance by biofeedback and question type (Experiment 1).

Table 2
Canonical correlations for affective states and metacognitive judgments predicting performance (Experiment 1).

Measure	Structure coefficients (r_s)					
	Accelerated		Baseline		Control	
	Function 1	Function 2	Function 1	Function 2	Function 1	Function 2
Predictor						
Text-based						
Arousal	**.324**	−.632	.178	.253	−.261	−.333
Valence	**.452**	−.627	−.254	.012	−.208	**.404**
EOL	**.421**	−.061	**.321**	.184	**.351**	−.260
JOL	**.368**	−.074	**.516**	.103	**.675**	−.351
RCJ	**.687**	.012	**.644**	−.598	**.512**	.198
Inference						
Arousal	.057	**.382**	**.322**	−.145	−.134	.104
Valence	−.339	**.497**	−.249	.012	.132	.242
EOL	.36	−.213	.102	**.307**	**.389**	.035
JOL	−.252	**.356**	.243	.201	−.603	−.168
RCJ	−.403	**.764**	−.404	**.389**	.109	−.155
Outcome						
Perf. (TB)	**.972**	.231	**.927**	−.383	**.861**	−.511
Perf. (INF)	.297	**.945**	.336	**.942**	**.546**	**.826**

Note: Structure coefficients (r_s) greater than 0.30 are in bold.

EOLs, JOLs, and RCJs, for both text based and inference questions) as predictors of the two learning outcome variables (mean learning outcomes for text based and inference questions) to evaluate the multivariate shared relationship between the two variable sets across biofeedback conditions (see Table 2). For an extensive review of how canonical correlations are conducted and interpreted, see Sherry and Henson (2005).

3.2.5.1. Accelerated condition. The analysis for the accelerated condition yielded two significant functions with squared canonical correlations (R_c^2) of 0.70 and 0.36 respectively. Collectively, the full model across both functions was statistically significant, Wilks' $\lambda = 0.19$, $F(20,76) = 4.82$, $p < .001$. Because Wilks' λ represents the variance unexplained by the model, using the equation $(1 - \lambda)$ yields the full model effect size as a r^2 metric. Thus, for the combined set of two canonical functions, the r^2 type effect size was 0.81, which indicates that the full model explained a substantial portion (81%) of the shared variance between the variable sets. Table 2 presents the standardized canonical correlations for Functions 1 and 2. The first column lists the *predictor variables* and *outcome variables* that were included in the analysis. Variables with correlations at and above 0.30 were determined to be relevant to individual functions and the overall model, since 0.30 is considered a moderate correlation (Cohen, 1988).

The pattern of correlations for Function 1 indicates that the relevant outcome variable was performance on text based questions, and all five text-based variables (EOL, JOL, RCJ, valence, arousal) were relevant predictors. All of these variables shared the same sign, indicating that they are all positively related. That is, increases in these variables were associated with increased performance on text based questions. Thus, Function 1 is indicative of affective states, metacognitive judgments, and performance outcomes that occur for text based questions.

For Function 2, we found that the relevant outcome variable was performance on inference questions, and the relevant predictor variables were JOLs, RCJs, and self-reported valence and arousal for inference questions. Similar to Function 1, these variables all shared a positive relationship with Function 2. Hence, the major difference between Functions 1 and 2 appears to be the lack of EOLs as a predictor for Function 2.

3.2.5.2. Baseline condition. Similar to the accelerated condition, the analysis for the baseline condition yielded two significant functions with squared canonical correlations of 0.52 and 0.31, respectively. The full model was statistically significant, Wilks' $\lambda = 0.33$, $F(20,76) = 2.80$, $p < .01$, with an r^2 effect size of 0.67.

The canonical correlations for Function 1 indicate that the relevant outcome variable was performance on text based questions, and the relevant predictor variables were EOLs, JOLs, and RCJs for text based questions. Thus, Function 1 seems to indicate that metacognitive judgments but not affective states are predictive of learning outcomes for text based questions. The relevant outcome variable for Function 2 was performance on inference questions, and the predictor variables were EOLs and RCJs for inference questions. Thus, these results indicate that affect was not a significant predictor of performance during baseline trials. Instead, only metacognitive judgments significantly predicted performance.

3.2.5.3. Control condition. The control condition yielded two functions with squared canonical correlations of 0.49 and 0.10, respectively. The full model was statistically significant using the Wilks' $\lambda = 0.47$ criterion, $F(20,76) = 1.77$, $p < .05$, with an r^2 type effect size of 0.54. Function 1 accounted for a moderate portion of the variance (0.52). However, Function 2 accounted for only a small portion of the variance (0.29), and was not statistically significant, $F(9,39) = 1.93$, $p > .05$. Thus, Function 2 will not be discussed.

The canonical correlations for Function 1 indicate that the relevant outcome variables were performance on text based and inference questions, and the relevant predictor variables were primarily EOLs, JOLs and RCJs for text-based questions, and EOLs for inference questions. Thus, the control condition is unique in that there is less of a distinction between text based and inference questions on Function 1. This finding is interesting because it indicates that when no biofeedback was presented the effect of question type on participants' affective states and metacognitive judgments was not as strong. More importantly, our findings show that affect did not appear to be a predictor of performance during control trials.

3.3. Discussion

This experiment was designed to test three hypotheses about the effect of biofeedback on participants' affective states, metacognitive judgments, and learning performance. In this section, we will describe how the findings relate to each of these hypotheses in turn.

Our findings provide moderate support for the *alert hypothesis*, compared to the *alarm hypothesis*. Specifically, the results indicated that when participants heard either accelerated *or* baseline false biofeedback, they experienced more positive/activating affective states compared to when they heard no biofeedback and experienced a positive/deactivating states (see Fig. 3). Compared to the no biofeedback control, participants made higher JOLs and RCJs, and achieved higher performance, when they received accelerated or baseline biofeedback. Interestingly, the effects of false biofeedback predominantly occurred when difficult inference questions had to be answered, thereby providing some support for the *question depth* hypothesis. There was no evidence for the *differential feedback hypothesis*. The general finding that affective states, metacognitive judgments, and performance during accelerated and baseline trials differed significantly from control trials suggests that it was the mere presence of any false biofeedback, rather than the *type* of false biofeedback, that caused these effects.

The canonical correlations across the three biofeedback conditions revealed some interesting findings about the relationship between the key dependent measures in this experiment. Of primary interest was the fact that the most complex relationships among key variables occurred when participants heard accelerated biofeedback. More specifically, although metacognitive judgments were strong predictors of performance across all types of biofeedback, affective states were only relevant predictor variables in the accelerated condition. This finding suggest that when participants were presented with an accelerated heart rate, their affective states were more likely to impact their performance on both text based and inference questions than when they were presented with a baseline or control heart rate. Importantly, the fact that the canonical correlation coefficients for participants' affective states and performance in the accelerated model were all positive indicates that affect was positively related to performance. Taken together, our findings suggest that the perception of physiological arousal can induce significant changes in learners' self-reported affective states, metacognitive judgments, and performance.

We were concerned that a potential limitation of Experiment 1 was that the Affect Grid, which measured participants' affective states, was presented at the end of each trial rather than immediately following the presentation of false biofeedback (i.e., immediately after studying the content slide and hearing an accelerated, baseline, or no heart rate). Because false biofeedback was manipulated in order to influence learners' affect, a more defensible position would be to require participants to complete the Affect Grid immediately after the presentation of false biofeedback. To address this issue, in Experiment 2 the Affect Grid was presented immediately following the presentation of false biofeedback, rather than at the end each trial.

In addition, in Experiment 1 we made the assumption that participants' physiological arousal aligned with the presentation of false biofeedback, and that participants appraised their physiological arousal as being related to some aspect of the learning task (i.e., *My heart is beating fast because I must not understand what I'm learning about*). Our manipulation check indicated that most participants did, in fact, believe the false biofeedback was indicative of their own physiological arousal. However, four participants *did not* believe they were hearing their own heart beat. This raised an interesting question: Could a person hearing a false heart rate, and *knowing* that it is false, still experience salient shifts in affective states and metacognitive judgments that ultimately impact performance? To address this question, Experiment 2 examined whether believing that false biofeedback is one's own heart beat is necessary for experiencing shifts in these processes, or if shifts arise even in the absence of such a belief.

4. Experiment 2

4.1. Method

4.1.1. Participants

Participants were 38 undergraduate students from a university in the US. The participants' mean age was 21.8 years (*SD* = 5.07), and there were 23 females (61%) in the sample. Most participants self-reported being Caucasian (56%) or African American (42%). All participants received $20 for participating in the experiment.

4.1.2. Design

We used a mixed 2 (Biofeedback: accelerated, baseline, control) × 2 (QT: text-based, inference) × 2 (Biofeedback Belief: self, other) design. In addition to the two within-subjects factors from Experiment 1 (Biofeedback and QT), Experiment 2 also included one between-subjects factor (Biofeedback Instruction: *self* vs. *recording*) (see Section 4.1.4 below). Participants were randomly assigned to either the *self* ($n = 19$) or *recording* ($n = 19$) condition.

4.1.3. Stimuli and software, apparatus, and material

The stimuli and software, apparatus, and materials were identical to Experiment 1. Similar to Experiment 1, type of false biofeedback (accelerated, baseline, or control) was presented randomly across trials.

4.1.4. Procedure

The experimental procedure was similar to Experiment 1, with two differences. First, as mentioned above, the Affect Grid was presented immediately following the presentation of false biofeedback, rather than at the end of each trial. Second, participants were randomly assigned to either the *self* condition or the *recording* condition. Participants in the *self* condition wore the Reebok™ heart rate monitor and were given the same instructions that were given to all participants in Experiment 1. Participants in the *recording* condition did not wear the monitor. These participants were informed that throughout the experiment they would hear a previously recorded, manually created auditory stimulus presented during particular trials. The main difference between the two conditions was that participants in the *self* condition presumably appraised the false biofeedback as being indicative of their own heart rate. On the other hand, participants in the *recording* condition presumably appraised the false biofeedback as being indicative of something or someone other than themselves (i.e., that the heart rate was previously recorded and had nothing to do with their own physiological arousal).

4.2. Results

We conducted a series of 3 × 2 × 2 (biofeedback × QT × Biofeedback Instruction) mixed ANOVAs and MANOVAs to examine the effect of these variables on participants' affective states, metacognitive judgments, and performance. We found no significant main effect or interactions involving Biofeedback Instruction in any of the analyses, indicating that believing that false biofeedback is indicative of one's own heart beat is not necessary in order to experience shifts in affective and metacognitive, processes and performance.

There was a concern that our sample size was not sufficiently large to detect significant differences between the *self* and *recording* conditions. In order to explore this possibility, we examined effect sizes associated with the main effect of biofeedback on key dependent variables. As Table 3 indicates, the effect sizes for most dependent variables were consistently small. Given these small effects, it is unlikely that power is an issue, because hundreds of par-

Table 3
Means, standard deviations, and effect sizes for main effects and interactions between conditions.

	Main effect (Condition)		
	Self *M* (*SD*)	Recording *M* (*SD*)	*d*
Affective			
Arousal	5.96 (.962)	5.63 (.592)	.413
Valence	5.24 (1.30)	5.42 (.913)	.160
Metacognitive			
EOL	4.18 (.833)	4.31 (.812)	.158
JOL	4.96 (.713)	4.81 (.610)	.226
RCJ	4.97 (.444)	4.97 (.573)	0.00
Cognitive			
Performance	.791 (.109)	.800 (.149)	.069

ticipants would be needed to detect these differences. For instance, a power analysis indicated that a sample of 1652 would be needed to detect the effect size of .069 for JOLs (power = .80, $\alpha = .05$, for a two-tailed independent samples *t*-test). Arousal was the only measure that yielded a small to medium effect size ($d = .41$). A post hoc power analysis indicated that achieved power to detect this .41 sigma effect with a two-tailed independent samples *t*-test was .24. This is well below the recommended value of 0.80 (Cohen, 1992).

With the exception of arousal, the *self* and *recording* conditions were incredibly similar on all other measures; thus, the subsequent discussion focuses only on the within-subjects effects of biofeedback and QT on the entire sample ($N = 38$). Descriptive data for these analyses is presented in Table 4, and replication patterns between the Experiment 1 and 2 are shown in Table 5.

4.2.1. Affective states

A 3×2 repeated measures MANOVA was conducted to examine the effect of false biofeedback on self-reported arousal and valence jointly. Similar to Experiment 1, this analysis revealed a significant main effect of feedback, Wilks' $\lambda = 0.81$, $F(4,146) = 4.12$, partial $\eta^2 = 0.10$, and a significant main effect of question type, Wilks' $\lambda = 0.68$, $F(2,36) = 6.76$, partial $\eta^2 = 0.32$.There was also a significant biofeedback × QT interaction, Wilks' $\lambda = 0.79$, $F(4,146) = 4.66$, partial $\eta^2 = 0.11$.

4.2.1.1. Arousal. Subsequent univariate analyses revealed that the main effect of biofeedback was replicated in Experiment 2 for participants' self-reported arousal, $F(2,36) = 6.85$, $MSE = 0.343$, partial $\eta^2 = 0.16$ (Accelerated = Baseline > Control). However, we found no significant main effect for QT, and no significant interaction. Rather that providing details on the post hoc tests, the major patterns in the data are summarized in Table 5. Patterns that replicated across experiments are presented in bold.

4.2.1.2. Valence. Similar to Experiment 1, we found no significant main effect for biofeedback on participants' self-reported valence. Consistent with Experiment 1, we found a significant main effect for QT, $F(1,37) = 17.2$, $MSE = 0.817$, partial $\eta^2 = 0.324$ (Text based > Inference), and a significant false biofeedback × QT interaction, $F(2,36) = 7.23$, $MSE = 0.614$, partial $\eta^2 = 0.157$.

4.2.1.3. Mapping valence and arousal on the Affect Grid. Similar to Experiment 1, we mapped participants' valence and arousal onto the Affect Grid to explore how they differed across accelerated, baseline, and no biofeedback (see Fig. 7). Consistent with Experiment 1, we found that participants' valence and arousal mapped onto the High Arousal/Positive Valence quadrant (Q1) when they received accelerated or baseline biofeedback, while mapping into the Low Arousal/Positive Valence quadrant (Q2) when they heard no biofeedback. After calculating each participants' Euclidean distance from neutral, we conducted a one-factor repeated measures ANOVA to determine if, like Experiment 1, Euclidean distance from neutral differed significantly across accelerated, baseline, and no biofeedback trials. Results of this analyses were significant, $F(2,36) = 8.23$, $MSE = 0.182$. A Bonferroni post hoc test revealed that participants' Euclidean distance from neutral was significantly greater for accelerated ($M = 1.54$, $SD = 0.858$, $d = 0.34$) and baseline trials ($M = 1.59$, $SD = 0.932$, $d = 0.38$) than control trials ($M = 1.24$, $SD = 0.928$). Thus, findings in Experiment 2 replicated findings from Experiment 1.

4.2.2. Metacognitive judgments

4.2.2.1. Ease of learning judgments. As in Experiment 1, we found a significant main effect for QT on participants' EOL judgments, $F(1,37) = 70.34$, $MSE = 0.186$, partial $\eta^2 = 0.655$ (Text Based > Inference). The main effect for biofeedback, and the biofeedback × QT interaction were not tested since EOLs occurred before the presentation of biofeedback.

4.2.2.2. Judgments of learning. Our findings for participants' JOLs were identical for our findings from Experiment 1. We failed to find a significant main effect for biofeedback, yet we did find a significant main effect for QT, $F(1,37) = 79.7$, $MSE = 0.284$, partial $\eta^2 = 0.68$, (Text Based > Inference), and a significant biofeedback × QT interaction, $F(2,36) = 7.87$, $MSE = 0.261$, partial $\eta^2 = 0.183$. Bonferroni post hoc analyses indicated that, in line with the overall trend from Experiment 1, biofeedback only had a significant effect on participants' JOLs when they were asked to answer an inference question, rather than a text based question (see Table 4).

Table 4
Descriptive statistics for dependent variables (Experiment 2).

	Text-based question			Inference question		
	Accelerated *M* (*SD*)	Baseline *M* (*SD*)	Control *M* (*SD*)	Accelerated *M* (*SD*)	Baseline *M* (*SD*)	Control *M* (*SD*)
Affective						
Arousal	5.30 (1.38)	5.15 (1.31)	5.14 (1.45)	5.47 (1.22)	5.35 (1.39)	4.95 (1.09)
Valence	5.84 (.991)	5.79 (1.04)	6.30 (1.10)	5.31 (1.16)	5.79 (.981)	5.34 (1.13)
Metacognitive						
EOL	4.74 (.889)	4.64 (.856)	5.03 (.759)	4.10 (.641)	3.93 (.929)	3.87 (.881)
JOL	5.16 (.871)	5.11 (.663)	5.37 (.564)	4.72 (.826)	4.67 (.864)	4.36 (.872)
RCJ	5.54 (.554)	5.29 (.534)	5.50 (.487)	4.61 (.884)	4.64 (.718)	4.32 (.864)
Cognitive						
Performance	.892 (.167)	.856 (.185)	.899 (.221)	.824 (.283)	.778 (.212)	.592 (.209)

Note: Performance scores indicate the proportion of correct answers out of all possible answers.

Table 5
Summary of replication patterns between Experiments 1 and 2.

Measure	Alert hypothesis	Alarm hypothesis	Main effect (Condition)		Condition × question type interaction			
					Text based		Inference	
			Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
Affective								
Valence	A > B = C	A < B = C	-	-	-	-	-	-
Arousal	**A > B = C**	**A > B = C**	**A = B > C**	**A = B > C**	**A = B = C**	**A = B = C**	**A = B > C**	**A = B > C**
Metacognitive								
JOL	A > B = C	A < B = C	-	-	C = B > A	A = C > B	A = B > C	A = B > C
RCJ	A > B = C	A < B = C	A > B > C	-	A = B = C	-	A > B > C	-
Cognitive								
Performance	A > B = C	A < B = C	A = B > C	**A = B > C**	**A = B = C**	**A = B = C**	**A = B > C**	**A = B > C**

Note: Ease of learning judgments were excluded from this table because they occurred prior to the presentation of false biofeedback.

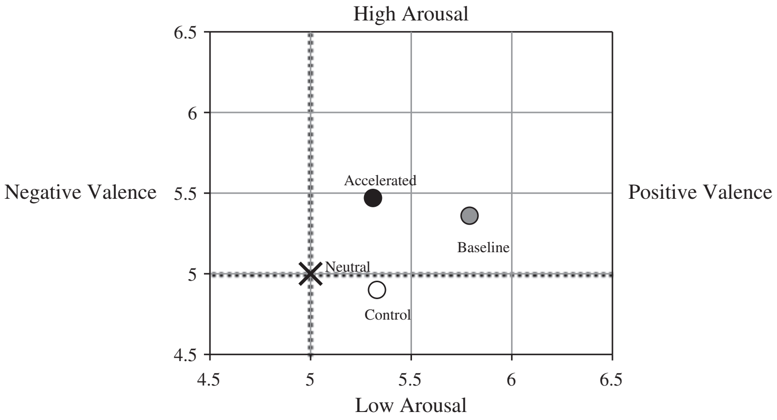


Fig. 7. Mean combined valence and arousal scores across biofeedback for inference questions (Experiment 2).

4.2.2.3. Retrospective confidence judgments.

Unlike Experiment 1, where we found a significant main effect for biofeedback on participants' RCJs, we found no significant main effect in Experiment 2. However, the main effect of QT remained significant, $F(1,37) = 97.8$, $MSE = 0.489$, partial $\eta^2 = 0.732$, (Text Based > Inference), as well as the biofeedback × QT interaction, $F(2,36) = 6.91$, $MSE = 0.191$, partial $\eta^2 = 0.16$. Once again, as Table 4 indicates, post hoc analyses revealed that the effect of false biofeedback was more prevalent for inference questions than text based questions. Specifically, although we found no significant effect of feedback on participants' RCJs for text based questions, we did find a significant effect for inference questions, revealing the following pattern Accelerated = Baseline > Control. This finding is a bit different from the pattern we found in Experiment 1 (Accelerated > Baseline > Control); however, RCJ's in the experimental conditions were greater than RCJ's in the control condition in both experiments.

4.2.3. Performance

Results from participants' scored responses on textbased and inference questions were identical to Experiment 1. We found a significant main effect for biofeedback, $F(2,36) = 7.54$, $MSE = 0.03$, partial $\eta^2 = 0.172$, (Accelerated = Baseline > Control). We also found a significant main effect for QT, $F(1,37) = 32.9$, $MSE = 0.04$, partial $\eta^2 = 0.474$, (Text Based > Inference) and a significant false biofeedback × QT interaction, $F(2,36) = 9.84$, $MSE = 0.041$, partial $\eta^2 = 0.212$. Similar to Experiment 1, Bonferroni post hoc analyses revealed that participants' performance only differed significantly for inference questions (Accelerated = Baseline > Control).

4.3. Discussion

In Experiment 2 we sought to replicate our findings for the *alert* vs. *alarm*, *differential biofeedback*, and *question depth* hypotheses. We also tested the *self* vs. *recording hypothesis* to determine if it was necessary for participants to believe that false biofeedback was indicative of their own physiological arousal in order to experience shifts in affective states, metacognitive judgments, and performance. Overall, similar to Experiment 1, we found moderate support for the *alert hypothesis*. Participants experienced more positive activating affective states and achieved higher learning performance when they were presented with false biofeedback than when they were presented with no biofeedback. Although we found no significant main effect of condition on participants' JOLs and RCJs, we found a significant condition × question type interaction, which revealed that the effect of condition was strongly influenced by the type of question participants were required to answer. Also similar to Experiment 1, we failed to find support for the *differential biofeedback* hypothesis, since there were generally no differences in affective states, metacognitive judgments, and performance between accelerated and baseline trials.

Further, we found support for the question depth hypothesis, since the effect of false biofeedback was most pronounced when participants were asked to answer challenging inference questions. Lastly, results indicated that the belief that false biofeedback is indicative of one's actual physiological arousal was *not* necessary in order for these shifts to occur. That is, the mere presentation of false biofeedback, regardless of participants' beliefs, was sufficient to impact affective states, metacognitive judgments, and performance outcomes. Thus, findings from Experiment 2 do not support the *self* vs. *recording* hypothesis. This finding warrants replication, however, because our sample was not sufficiently large to detect significant differences between conditions. We are confident that this is not a critical issue due to the relatively small effect sizes between conditions on all measures except arousal. There was a medium effect between conditions for arousal, but we lacked adequate power to detect this difference, so we cannot be sure that participants in the *self* and *recording* conditions did not significantly differ in their levels of arousal.

We also found that several key patterns from Experiment 1 replicated in Experiment 2 (see Table 5). Specifically, we found that biofeedback had a significant impact on learners' self-reported arousal, and yet, like Experiment 1, no significant impact on self-reported valence. Although we found some small differences between Experiments 1 and 2 in the effect of biofeedback on participants' metacognitive judgments, many results replicated perfectly between the two experiments. Similar to Experiment 1, participants made significantly higher EOLs, JOLs, and RCJs when they were asked to answer a text-based question rather than an inference question. We also found that the effect of biofeedback on participants' metacognitive judgments (JOLs and RCJs) was most pronounced for inference questions than text based questions.

We found that the main effect of biofeedback, and the biofeedback × QT interaction replicated perfectly for participants' performance between Experiments 1 and 2. Specifically, we found that for inference questions participants performed significantly better when they received accelerated or baseline biofeedback than when they received no biofeedback, and that the effect of biofeedback was more pronounced for difficult inference questions than text based questions.

Canonical correlation analyses were not conducted on data from Experiment 2 due to the relatively small sample size ($n = 19$ per condition) and large set of independent variables (10 overall), which would not provide enough power for meaningfully interpreting the results. However, we feel that this is not a major limitation because the goal of Experiment 2 was not to replicate Experiment 1. Rather, the goal of Experiment 2 was to determine whether beliefs about the relationship between false biofeedback and actual physiological arousal influenced the effect of false biofeedback on participants' affective states, metacognitive judgments, and performance.

5. General discussion

In this section we discuss the extent to which the experimental findings supported our hypotheses, discuss the theoretical and practical implications of our findings, and present limitations and future directions of this research.

5.1. Hypotheses, major findings, and implications

We examined the effect of false biofeedback and question type on participants' affective states, metacognitive judgments, and performance during learning with multimedia. We tested several hypotheses about the effect of biofeedback on each of these constructs. The *alarm hypothesis* predicted that false biofeedback would lead participants to report high arousal and negative valence, make significantly less confident metacognitive judgments, and achieve lower learning performance. Conversely, the *alert hypothesis* predicted that participants would report high arousal and positive valence, make significantly more confident metacognitive judgments, and achieve higher learning performance when they were presented with false biofeedback. The *differential feedback hypothesis* predicted that accelerated and baseline biofeedback would have differential effects on learners' affective states and metacognitive judgments and learning performance. The question depth hypothesis predicted that the effect of false biofeedback would be most prominent for inference questions, since these questions are more challenging than text based questions. Also, in Experiment 2, the self vs. recording hypothesis examined whether participants needed to believe that the false biofeedback was indicative of their own physiological arousal in order to experience shifts in affective states, metacognitive judgments, and performance.

The results from Experiments 1 and 2 supported the *alert hypothesis* instead of the *alarm hypothesis*. Findings across both experiments indicate that false biofeedback differed significantly from the no biofeedback control, and was associated with High Arousal/Positive Valence states, more confident metacognitive judgments, and increased performance, compared to no biofeedback control. We found no support for the *differential biofeedback hypothesis,* since there were generally no significant differences in participants' affective states, metacognitive judgments, and performance between accelerated and baseline biofeedback trials. Findings across both experiments supported the *question depth hypothesis*, since the effects of false biofeedback were more prevalent when inference questions had to be answered.

Lastly, in Experiment 2 we found no support for the *self* vs. *recording* hypothesis, indicating that participants did not need to believe that the false biofeedback was indicative of their own physiological states in order to experience shifts in their affective states, metacognitive judgments, and performance. As discussed in the Introduction, it appears that these findings offer support for the occurrence of physiological alignment in the presence of false biofeedback. That is, it is perhaps likely that even when participants were explicitly instructed that the false biofeedback was not indicative of their physiological arousal, participants still may have experienced physiological alignment with the feedback, which may have been responsible for the shifts in affect, metacognition, and performance. An alternative position is that there may have been differences between the *self* vs. *other* condition for arousal, but that issues of power prevented us from detecting these differences. It is impossible to explore this hypothesis because we did not collect participants' actual physiological arousal. This is a limitation that will be subsequently discussed.

Overall, our findings suggest that false biofeedback is an effective method for manipulating affective states and metacognitive judgments, and positively impacting learning performance, and that perhaps the effects of false biofeedback are due primarily to the *increased monitoring* caused by the presence of the stimulus, rather than the type of stimulus. An alternative position is that perhaps accelerated biofeedback and baseline biofeedback both had facilitating effects on affect, metacognition, and performance, but for different reasons. For example, results indicated that participants reported increased arousal in both accelerated and baseline trials. Perhaps accelerated biofeedback increased arousal, which was associated with feelings of engagement and improved performance, but was also distracting and took participants' attention slightly away from the learning task. On the other hand, baseline biofeedback might have increased arousal to a lesser extent, but because baseline feedback is less distracting than accelerated feedback (since it is more akin to a regular heart rate), participants might have more cognitive resources available to monitor their learning and focus on the task.

An interesting topic for future research is to explore this question further in order to determine what level of arousal is most beneficial to learning, and what level of arousal is harmful. Yerkes and Dodson (1908) concluded that arousal and performance are related via a task-moderated inverted-U curve, meaning that arousal will generally facilitate performance until it has reached some unknown threshold, at which point the arousal will become detrimental to performance. Current research on boredom and anxiety in learning contexts offers support for this claim, since boredom (a state in which arousal is too low) and anxiety (a state in which arousal is too high) are both associated with decrements in learning (Daniels, Stupnisky, et al., 2009; Daniels, Pekrun, et al., 2009; Goetz et al., 2007; Harris, 2000; Hembree, 1988; Kass, Vodanovich, Stanny, & Taylor, 2001; Pekrun et al., 2009; Zeidner, 2007). It is possible that *flow*, a state of intense engagement and concentration, might occur when there is an optimal level of arousal, caused by an appropriate balance between skills and challenge (Csikszentmihalyi, 1990, 1997; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Shernoff, Csikszentmihalyi, Schneider, & Scernoff, 2003). However, to our knowledge, researchers have yet to identify the precise level of physiological or subjective arousal that is optimal for learning, which is an issue that clearly deserves future exploration.

Our findings also demonstrated that there are clear, identifiable relationships between affect, metacognition, and performance, but only for inference questions that require deep reasoning and thought. Most existing theories on affect and learning (Boekaerts, 2001, 2011; Boekaerts et al., 2000; Pekrun et al., 2009) recognize that affect plays an important role in the process of learning. However, many of these theories could be expanded to provide a clearer picture of how affect influences self-regulation, behavior, and learning performance. For example, existing theories and models could be improved by offering an explanation of how affect relates to other key processes that occur during learning, including its effect on: (1) integration of new information with prior knowledge, (2) allocation of attentional resources to the learning task, (3) metacognitive judgments regarding the perceived difficulty of the task, learners' emerging understanding, and their confidence in their responses, (4) the appropriate (or inappropriate) allocation of study-time, and (5) learning performance. Although the present research does not address all of these issues, it does highlight some of the important interactions between affective states, metacognitive judgments, and performance during learning. Future research directed toward refining and expanding existing theories of learning with multimedia (e.g., Mayer, 2009) should focus on determining if these results replicate across different topics (e.g., ecology, computer literacy), student populations (such as middle- and high-school students or adult learners), and learning tasks.

5.2. Educational implications

Our findings have a number of practical implications because they suggest that learners' affective states can significantly hinder or facilitate performance during science learning. Science learning is particularly challenging because learners are required to comprehend intricate, multi-level systems of relationships between concepts that are often novel and unfamiliar (Azevedo, 2009). Science learning with computerized environments, which requires learners to integrate multiple instances of text and diagrams and navigate hundreds of links, can make the learning experience even more challenging, and learners may experience a vast array of affective states during such learning episodes. These affective experiences can be even more pronounced when the learner has low domain knowledge about the topic, lacks interest, or has a predisposition towards test anxiety. Further, girls may tend to experience more negative affect during computerized science learning, since research has demonstrated that females tend to struggle with STEM (Science, Technology, Engineering, and Mathematics) topics more than males and tend to have negative attitudes toward science (Jenkins & Nelson, 2005).

As such, it is important that our results serve to improve the design of computer-based learning environments such as multimedia, hypermedia, and intelligent tutoring systems (ITSs) that are intended to teach learners about difficult (and often anxiety-provoking) science and math topics (Azevedo, Behnagh, Duffy, Harley, & Trevors, 2012). There is a need for computerized learning environments to be sensitive to the dynamic and complex processes that occur during STEM learning in order to help learners acquire knowledge in a manner that effectively coordinates cognition and affect (D'Mello & Graesser, 2010; D'Mello, Picard, & Graesser, 2007). For example, ITSs that use pedagogical agents to scaffold learners' understanding of complex science topics might benefit from the use of physiological and bodily measures that can detect shifts in learners' affective states in real-time. If a learner shifts to a negative affective state (i.e., stress, boredom), a system that is sensitive to these shifts could help learners: (1) become aware of such states, and (2) transition out of these affective states by modeling, prompting, and scaffolding appropriate self-regulatory processes.

How can computerized learning environments accurately record and assess learners' affective states? Current research is being conducted to determine if affect sensors embedded in ITSs can detect the affective states that occur during learning (Conati & Maclaren, 2009; D'Mello, Craig, Sullins, & Graesser, 2006; D'Mello et al., 2007; Forbes-Riley, Rotaru, & Litman, 2008; McQuiggan, Mott, & Lester, 2008; Woolf et al., 2009). For example, D'Mello and colleagues have developed an ITS that detects and responds to boredom, confusion, and frustration by monitoring contextual cues, facial expressions, and body movements (posture) (D'Mello & Graesser, 2010). Importantly, this affect-sensitive ITS yielded impressive learning gains for struggling learners with low domain knowledge when compared to a version of the ITS that was sensitive to learners' cognitive states but not their affective states.

There are still many unanswered questions regarding affect during computerized learning. Once learner affect can be automatically detected and assessed in situ during learning, how can ITSs help learners monitor and regulate their affective states while they learn? More importantly, can recognizing and regulating their affect help learners become more metacognitively aware of their emerging understanding of the topic, which may lead to more meaningful learning? Answers to these questions would enhance advances that have already been made in the area of cognitive and metacognitive regulatory processes during computerized science learning (e.g., Aleven, Roll, McLaren, & Koedinger, 2010; Azevedo & Strain, 2011; Azevedo et al., 2010).

Although answers to the questions raised above will require further research and technological development, some relevant insights can be gleaned from the present research. We have demonstrated that there appear to be some merits to affect induction during learning, and that the method of false biofeedback may be a useful tool for inducing affective states during short (30 min to 1 h) learning episodes. For example, it might be the case that learners who are disinterested, unmotivated, or bored (states that usually are accompanied by low physiological arousal, negative affect, and poor learning) (Pekrun et al., 2010), may profit from the presentation of false biofeedback (or a similar arousing stimulus) during learning. Our findings demonstrate that false biofeedback is associated with more positively valenced affect and higher arousal, which in turn leads to more confident metacognitive judgments and increased learning. Our findings demonstrate that false biofeedback may influence these effects even in learners who *know* that the false heart rate is not indicative of their own physiological arousal. This is a critical point because it implies that false biofeedback is a method of inducing affect which may lead to improved learning performance without having to deceive learners (i.e., by telling learners that they are hearing their own heart rate when they are not), and without expensive equipment, or extensive training on various methods to induce affective states.

5.3. Limitations and future directions

This research was limited by the typical challenges of studying affect during learning. The biggest of these challenges is the fact that affect is a complex construct with murky boundaries and substantial individual differences in expression and experience (Gross & Barrett, 2011; Izard, 2007). In these experiments, abrupt changes in participants' affective state during a single trial may have altered their metacognitive judgments, which could have had a significant impact on their performance. Our inability to detect these within-trial shifts may have limited our ability to make process-pure inferences regarding the effect of false biofeedback on metacognition and performance. Further, both experiments were limited by their use of only one assessment of learners' affective states during the learning session. The assessment of learners' affect could have been improved if we had collected self-reports on discrete affective states such as boredom, engagement, frustration, confusion throughout the session.

The method of inducing affective states was also associated with some of its own challenges. One major limitation of the methodology was our inability to ensure that induced affective states from one trial did not persevere into the following trial. We attempted to control for such carry-over effects by presenting many tasks in between the presentation of biofeedback in one trial and the beginning of the next trial. However, we cannot be completely confident that such effects did not occur.

Another possible limitation is that physiological responses (i.e., heart rate and skin conductance) were not collected. Previous studies that have used false biofeedback methodologies suggest that relying on self-reports (rather than self-reports *and* physiological responses) is a defensible method, because (a) physiological arousal has been shown to align with false biofeedback (Ehlers et al., 1988; Holroyd et al., 1984; Lichstein & Hoelscher, 1989), and (b) self-reports of affective experiences tend to correlate with false biofeedback and actual physiological arousal (Barefoot & Straub, 1971; Gatchel, Korman, Weiss, Smith, & Lewis, 1978; Harris & Jellison, 1971; Kirsch & Lynn, 1999; Ma-Kellams, Blascovich, & McCall, 2012; McKinney & Gatchel, 1982; Palace, 1995; Valins, 1966). In fact, following the publication of several empirical studies in the 1970s that reported the strong correlation among false biofeedback, actual physiological arousal, and self-report affective experiences, Harris and Katkin (1975) concluded that this link is sufficiently strong that researchers could consider using false biofeedback without evaluating physiological arousal.

Although previous research offers support for not measuring physiological arousal, we acknowledge that results from this study would be much more interpretable if physiological arousal had been measured. Collecting participants' physiological responses would have been useful for determining: (1) if physiological arousal did, in fact, align with the type of false biofeedback, and (2) whether affective states persevered between trials (for example, increased physiological arousal in accelerated trials was maintained at the beginning on baseline trials).

Another limitation is that affective states that are experimentally induced and occur in a laboratory setting may be qualitatively different than those that occur naturally during studying or test-taking. Finding ways to induce affective states that are similar to those that occur naturally, or finding ways to examine naturally occurring affective states with fine-grained temporal resolutions without interfering or altering the affective states themselves, would dramatically improve the interpretability of the results from this experiment. This is an important item for future work.

Lastly, and perhaps most importantly for future research, we cannot determine from our findings whether the appraisal of false biofeedback is causally linked to changes in participants' affective states or if these changes are the result of the simple alignment of actual physiological arousal with false heart rates. The appraisal theoretic view we adopted posits that the experience of affect is inextricably linked to the types of appraisals made about situations in the environment (Clore & Ortony, 2010; Lazarus, Kanner, & Folkman, 1980; Roseman, 1984; Schachter & Singer, 1962; Smith & Ellsworth, 1985). As such, appraisal theorists might suggest that the findings of the current experiments were attributable to the appraisals that participants made when they were presented with false biofeedback. Somatic theories, on the other hand, posit that affective states are reflex-like experiences that can occur in the absence of cognitive appraisal and are the result of unique patterns of physiological responding (Cacioppo, Bernston, & Klein, 1992; Levinson, Ekman, & Friesen, 1990; Panksepp, 1998). According to this view, the specific physiological response *is* the affective experience, even in the absence of appraisal processes (Cacioppo, Bernston, Larsen, Poehlmann, & Ito, 2000; James, 1884). The somatic marker hypothesis (Damasio, 1994) posits that *primary inducers* are innate or learned stimuli that cause pleasant or aversive states that automatically elicit a somatic response and subsequent affective experience. In essence, this automatic process can short-circuit the need for appraisal processes in order for an affective experience to occur. Somatic theorists might predict that appraisal did not play a critical role in the distinct shifts in affect, metacognitive judgments, and performance in the present experiments. Instead, it is perhaps the case that false heart rates served as primary inducers that elicited automatic somatic and affective responses. Exploring these contrasting hypotheses in the current experiments is confounded by the fact that we did not ask participants to report the appraisals they made about the false biofeedback, and did not collect participants' physiological responses. As such, further research is needed in order to deepen our understanding of the links between affect, metacognition, and performance during learning.

6. Conclusion

There is a need for more empirically-driven research directed toward understanding of the relationship among affect, metacognition and performance during multimedia learning. As theoretical and conceptual issues are resolved and methodological techniques are improved, the elusive role of affect may be disambiguated, leading researchers to more fully understand the consequences of affect on learning, and to develop interventions and computerized environments that effectively coordinate learners' affective, metacognitive, and cognitive states during learning.

Acknowledgments

This research was supported by the National Science Foundation (NSF) (DRL 0633918; IIS 0841835; DRL 1008282), and the Social Sciences and Humanities Research Council of Canada, awarded to the second author, and NSF (HCC 0834847, DRL 1108845) awarded to the third author. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of these funding sources.

References

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*, 224–233.

Azevedo, R. (2009). Theoretical, methodological, and analytical challenges in the research on metacognition and self-regulation: A commentary. *Metacognition and Learning, 4*, 87–95.

Azevedo, R., Behnagh, R., Duffy, M., Harley, J., & Trevors, G. (2012). Metacognition and self-regulated learning in student-centered leaning environments. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of student-center learning environments* (2nd ed., pp. 171–197). New York: Routledge.

Azevedo, R., Moos, D., Johnson, A. M., & Chauncey, A. (2010). Measuring cognitive and metacognitive regulatory processes used during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*(4), 213–223.

Azevedo, R., & Strain, A. C. (2011). Integrating cognitive, metacognitive, and affective regulatory processes with MetaTutor. In R. Calvo & S. K. D'Mello (Eds.), *New perspectives on affect and learning technologies*. New York, NY: Springer.

Azevedo, R., & Witherspoon, A. M. (2009). Self-regulated learning with hypermedia. In A. Graesser, J. Dunlosky, & D. Hacker (Eds.), *Handbook of metacognition in education* (pp. 319–339). Mahwah, NJ: Erlbaum.

Baker, R., D'Mello, S., Rodrigo, M., & Graesser, A. (2010). Better to be frustrated than bored: The incidence and persistence of affect during interactions with three different computer-based learning environments. *International Journal of Human–Computer Studies, 68*(4), 223–241.

Barefoot, J. C., & Straub, R. B. (1971). Opportunity for information search and the effect of false heart rate feedback. *Journal of Personality and Social Psychology, 17*(2), 154–157.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion, 23*(7), 1284–1306.

Barrett, L., Mesquita, B., Ochsner, K., & Gross, J. (2007). The experience of emotion. *Annual Review of Psychology, 58*, 373–403.

Bless, H. (2000). The interplay of affect and cognition: The mediating role of general knowledge structures. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 201–222). New York, NY: Cambridge University Press.

Boekaerts, M., Pintrich, P., & Zeidner, M. (Eds.). (2000). *Handbook of self-regulation*. San Diego, CA: Academic Press.

Boekaerts, M. (2001). Context sensitivity: Activated motivational beliefs, current concerns and emotional arousal. In S. Volet & S. Jarvela (Eds.), *Motivation in learning contexts: Theoretical and methodological implications* (pp. 17–31). Elmsford, NY: Pergamon Press.

Boekaerts, M. (2009). Goal-directed behavior in the classroom. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 105–122). New York, NY: Routledge.

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408–425). New York: Routledge.

Bower, G. H., & Forgas, J. P. (2000). Affect, memory, and social cognition. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, & P. M. Niedenthal (Eds.), *Cognition and emotion* (pp. 87–168). New York: Oxford University Press.

Braten, I., Samuelstuen, M. S., & Stromso, H. I. (2004). Do students' self-efficacy beliefs moderate the effects of performance goals on self-regulatory strategy use? *Educational Psychology, 24*, 231–247.

Brosch, T., Pourtios, G., & Sander, D. (2010). The perception and categorization of emotional stimuli: A review. *Cognition and Emotion, 24*, 377–400.

Burkett, C., & Azevedo, R. (2012). The effect of multimedia discrepancies on metacognitive judgments. *Computers in Human Behavior, 28*, 1276–1285.

Cacioppo, J. T., Bernston, G. G., & Klein, D. J. (1992). What is an emotion? The role of somatovisceral afference, with special emphasis on somatovisceral "illusions". *Review of Personality and Social Psychology, 14*, 63–98.

Cacioppo, J. T., Bernston, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In R. Lewis & J. M. Haviland-Jones (Eds.), *The handbook of emotion* (pp. 173–191). Guilford Press.

Cacioppo, J. T., Klein, D. J., Berntson, G. G., & Hatfield, E. (1993). The psychophysiology of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 143–154). New York: Guilford Press.

Calvo, R., & D'Mello, S. (Eds.). (2011). *New perspectives on affect and learning technologies*. New York: Springer.

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*, 43–54.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393–399.

Clore, G. L. (1992). Cognitive phenomenology: Feeling and the construction of judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 133–164). Hillsdale, NJ: Erlbaum.

Clore, G. L., & Ortony, A. (2010). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628–644). New York, NY: Guilford Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*, 267–303.

Craig, S., Graesser, A., Sullins, J., & Gholson, J. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29*, 241–250.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New Tork: Harper-Row.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life. The mastermind series*. New York: Basic Books.

Damasio, A. R. (1994). *Descarte's error: Emotion, reason, and the human brain*. New York: Grosset/Putnam.

Daniels, L. M., Pekrun, R., Stupnisky, R. H., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology, 101*(4), 948–963.

Daniels, L. M., Stupnisky, R. H., Pekrun, R., Haynes, T. L., Perry, R. P., & Newell, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology, 101*(4), 948–963.

D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education, 16*, 3–28.

D'Mello, S. K., & Graesser, A. C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*(2), 147–187.

D'Mello, S., & Graesser, A. (2011a). The half-life of cognitive–affective states during complex learning. *Cognition and Emotion, 25*(7), 1299–1308.

D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*, 145–157.

D'Mello, S., & Graesser, A. (2011b). Emotions during learning with AutoTutor. In P. Durlach & A. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 117–139). Cambridge, UK: Cambridge University Press.

D'Mello, S., Lehman, B., & Person, N. (2011). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education, 20*(4), 361–389.

D'Mello, S. K., Picard, R., & Graesser, A. C. (2007). Towards and affect sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53–61.

Duffy, E. (1962). *Activation and behavior*. New York: Wiley.

Dunlosky, J., & Theide, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*, 779–788.

Dunosky, J., & Metcalfe, J. (2009). *Metacognition: A textbook for cognitive, educational, life span and applied psychology*. Newbury Park, CA: Sage.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*, 3–14.

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

Ehlers, A., Margraf, J., Roth, W. T., Taylor, C. B., & Birbaumer, N. (1988). Anxiety induced by false heart rate feedback in patients with panic disorder. *Behavior Research and Therapy, 26*, 1–11.

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99*, 550–553.

Elliot, A. J., & Thrash, T. M. (2002). Approach–avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology, 82*, 804–818.

Eysenck, M. W., Derekshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336–353.

Fielder, K. (2001). Affective states trigger processes of assimilation and accommodation. In K. Martin & G. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 85–98). Mahwah, NJ: Erlbaum.

Forbes-Riley, K., Rotaru, M., & Litman, D. (2008). The relative impact of student affect on performance models in a spoke dialogue intelligent tutoring system. *User modeling and User-Adapted Interaction, 18*, 11–43.

Fredrickson, B. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218–226.

Fredrickson, B., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought–action repertoires. *Cognition and Emotion, 19*(3), 313–332.

Frijda, N. H. (2009). Emotions, individual differences and the time course: Reflections. *Cognition and Emotion, 23*(7), 1444–1462.

Gatchel, R. J., Korman, M., Weiss, Charles. B., Smith, D., & Lewis, C. (1978). A multiple-response evaluation of EMG biofeedback performance during training and stress-induction conditions. *Psychophysiology, 15*(3), 253–258.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N., Nathan, C., & Ludke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99*(4), 715–733.

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*(1), 104–137.

Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web Tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*, 89–105.

Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*, 8–16.

Hacker, D., Dunlosky, J., & Graesser, A. C. (2009). *Handbook of metacognition in education*. New York: Routledge.

Harachiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*, 638–645.

Harris, M. B. (2000). Correlates and characteristics of boredom proneness and boredom. *Journal of Applied Social Psychology, 30*, 576–598.

Harris, V. A., & Jellison, J. M. (1971). Fear-arousing communications, false physiological feedback, and the acceptance of recommendations. *Journal of Experimental Social Psychology, 7*(3), 269–279.

Harris, V. A., & Katkin, E. S. (1975). Primary and secondary emotional behavior: An analysis of the role of autonomic feedback on affect, arousal, and attribution. *Psychological Bulletin, 82*(6), 904–916.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*, 47–77.

Holroyd, K. A., Penzien, D. B., Hursey, K. G., Tobin, D. L., Rogers, L., Holm, J. E., et al. (1984). Change mechanisms in EMG biofeedback training: Cognitive changes underlying improvements in tension headache. *Journal of Consulting and Clinical Psychology, 52*, 1039–1053.

Isen, A. (2001). An influence of positive affect on decision making in complex situations: Theoretical issues and practical implications. *Journal of Consumer Psychology, 11*, 75–85.

Isen, A. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. Haviland-Jones, & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548–573). New York, NY: Guilford.

Isen, A. (2010). Some ways in which positive affect influences decision making. In M. Lewis, J. M. Haviland-Jones, & L. Feldman-Barrett (Eds.), *Handbook of emotions* (pp. 548–573). New York, NY: Guilford Press.

Isen, A. (2004). Some perspectives on positive feelings and emotions: Positive affect facilitates thinking and problem solving. In A. S. R. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium, studies in emotion and social interaction* (pp. 263–281). New York, NY: Cambridge University Press.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2*(2), 260–280.

James, W. (1884). What is an emotion? *Mind, 9*, 188–205.

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Educational Psychology, 134*, 308–326.

Jenkins, E. W., & Nelson, N. W. (2005). Important but not for me: Students' attitudes toward secondary school science in England. *Research in Science and Technological Education, 23*, 41–57.

Karsdorp, P. A., Kindt, M., Rietveld, S., Everaerd, W., & Mulder, B. (2009). False heart rate feedback and the perception of heart symptoms in patients with congenital heart disease and anxiety. *International Journal of Behavioral Medicine, 16*(1), 8–81.

Kass, S. J., Vodanovich, S. J., Stanny, C. J., & Taylor, T. M. (2001). Watching the clock: Boredom and vigilance performance. *Perceptual and Motor Skills, 92*, 969–976.

Kirsch, I., & Lynn, S. J. (1999). Automaticity in clinical psychology. *The American Psychologist, 54*, 504–515.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology, 135*, 36–69.

Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lazarus, R. S., Kanner, A. D., & Folkman, S. (1980). Emotions: A cognitive-phenomenological analysis. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience. Theories of emotion* (Vol. I, pp. 189–217). New York: Academic Press.

Leonesio, R. J., & Nelson, T. O. (1990). Do different measures of metamemory tap the same underlying aspects of memory? *Journal of Experimental Psychology, 16*, 464–470.

Levinson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*(4), 363–384.

Lichstein, K. L., & Hoelscher, T. J. (1989). Biofeedback to enhance psychophysiological effects of imagery. *Behaviour Research and Therapy, 27*, 569–572.

Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review, 18*(4), 307–314.

Linnenbrink, E. A., & Pintrich, P. R. (2002). The role of motivational beliefs in conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 115–235). Dordretch, Netherlands: Kluwer Academic Publishers.

Linninbrink, L. A. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Amstedam: Elsevier.

Ma-Kellams, C., Blascovich, J., & McCall, C. (2012). Culture and the body: East–west differences in visceral perception. *Journal of Personality and Social Psychology, 102*, 718–728.

Mandler, G. (1975). *Mind and emotion*. New York: Wiley.

Mandler, G. (1999). Emotion. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive science. Handbook of perception and cognition* (2nd ed., pp. 367–382). San Diego, CA: Academic Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

McKinney, M. E., & Gatchel, R. J. (1982). The comparative effectiveness of heart rate biofeedback, speech skills training, and a combination of both in treating public-speaking anxiety. *Biofeedback and Self-regulation, 7*(1), 71–87.

McQuiggan, S. B., Mott, B., & Lester, J. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction, 18*, 81–123.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*, 349–363.

Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377–390.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York: Academic Press.

Newcombe, N. S., Ambady, N., Eccles, J., Gomez, L., Klahr, D., Linn, M., et al. (2009). Psychology's role in mathematics and science education. *American Psychologist, 64*, 538–550.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. New York: Cambridge University Press.

Palace, E. M. (1995). Modifcation of dysfunctional patterns of sexual response through autonomic arousal and false physiological feedback. *Journal of Consulting and Clinical Psychology, 63*(4), 604–615.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Pekrun, R. (2010). Academic emotions. In T. Urdan (Ed.). *APA educational psychology handbook* (Vol. 2). Washington, DC: American Psychological Association.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115–135.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 9–32). San Diego: Elsevier.

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

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P., & Schunk, D. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle, NJ: Prentice-Hall, Inc.

Programme for International Student Assessment (2009). *Learning trends: Changes in student performance since 2000.* http://dx.doi.org/10.1787/9789264091580-en.

Roseman, I. (1984). Cognitive determinants o f emotions: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Emotions, relationships, and health* (Vol. 5, pp. 11–36). Beverly Hills: Sage.

Rosenberg, E. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*(3), 247–270. http://dx.doi.org/10.1037//1089-2680.2.3.247.

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*, 493–502.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*(5), 379–399.

Schunk, D. (2001). Social-cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 125–152). Mahwah: NJ: Erlbaum.

Schutz, P., & Pekrun, R. (Eds.). (2007). *Emotion in education*. San Diego, CA: Academic Press.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion, 14*, 433–440.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry, 14*, 296–303.

Schwarz, N., & Skurnik, I. (2003). Feeling and thinking: Implications for problem solving. In J. Davidson & R. Sternberg (Eds.), *The nature of problem solving* (pp. 263–292). Cambridge, England: Cambridge University Press.

Schwarz, N. (2011). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 289–308). Los Angeles, CA: Sage.

Shapiro, A. M. (2008). Hypermedia design as learner scaffolding. *Education Technology Research, 56*, 29–44.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Scernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*, 158–176.

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting Canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality and Assessment, 84*(1), 37–48.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813–838.

Stein, N., Hernandez, M., & Trabasso, T. (2008). Advances in modeling emotions and thought: The importance of developmental, online, and multilevel analysis. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 574–586). New York: Guilford Press.

Stein, N. L., & Levine, L. J. (1991). Making sense out of emotion. In W. Kessen, A. Ortony, & F. Kraik (Eds.), *Memories, thoughts, and emotions: Essays in honor of George Mandler* (pp. 295–322). Hillsdale: Earlbaum.

Valins, S. (1966). Cognitive effects of false heart rate biofeedback. *Journal of Personality and Social Psychology, 4*, 400–408.

Wigfield, A., Eccles, J. S., Schiefele, U., Roesner, R., & Davis-Kean (2006). Development of achievement motivation. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (pp. 933–1002). Hoboken, NJ: John Wiley.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York, NY: Taylor & Francis.

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–32). New York: Routledge.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology, 3*, 129–164.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp. 165–184). San Diego, CA: Academic Press.

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*, 64–70.

Zimmerman, B., & Schunk, D. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York: Routledge.