The Pennsylvania State University

The Graduate School

Department of Information Sciences and Technology

IMPACTS OF USER SENTIMENT ON INFORMATION RECALL, INTRINSIC MOTIVATION, AND ENGAGEMENT IN THE CONTEXT OF INTELLIGENT TUTORING SYSTEMS

A Thesis in

Information Sciences and Technology

by

Luke R. Metaxas

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2018

The thesis of Luke R. Metaxas was reviewed and approved by the following:

Frank E. Ritter
Professor of Information Sciences and Technology
Thesis Advisor

Edward J. Glantz
Associate Teaching Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang
Associate Professor of Information Sciences and Technology

Korey MacDougall
Postdoctoral Scholar and Special Signatory

Andrea H. Tapia
Associate Professor of Information Sciences and Technology
Director of Graduate Programs

*Signatures are on file in the Graduate School

ABSTRACT

Socrates once stated: “Education is the kindling of a flame, not the filling of a vessel.” Intelligent tutoring systems use computational models to efficiently “fill the vessel.” However, research is limited on how these systems can enable “kindling of a flame.” To explore motivation-adaptive learning, this study assesses the learning impact of an individual’s sentiments (emotional associations) on three learning outcomes: information recall, intrinsic motivation, and engagement. Seventy volunteers took two computer-based tutors and provided self-report measures throughout their learning. The learning impacts of topic sentiment and learning-medium sentiment were measured separately and compared. For both topic and learning-medium sentiment, results showed positive linear relationships between net sentiments and intrinsic motivation and net sentiments and engagement. A negative linear relationship between negative sentiments and information recall was also identified. Findings were summarized, and four computer-based instruction design recommendations were provided. Recommendations include: to use learner emotion data in the form of sentiments to better understand learning outcomes, to align sentiment measurement strategies with the tutor’s purpose, to account for the impact of prior sentiments, and to monitor sentiment change.

TABLE OF CONTENTS

LIST OF FIGURES vii

LIST OF TABLES viii

ACKNOWLEDGEMENTS ix

Chapter 1: Introduction 10

Chapter 2: Literature Review 3

- Intelligent Tutoring Systems (ITSs) 3
  - Effectiveness of CBI and ITSs 3
  - ITS Components 4
  - The Learner Model and ITS Standardization 6
- Sentiment 10
  - Sentiment Overview 11
  - Distinguishing Between Related Terms 14
- Engagement and Flow 17
  - Flow Theory Overview 18
  - Flow and Education 19
- Intrinsic Motivation and Internalization 20
  - Self-Determination Theory Overview 21
  - The Self-Determination Continuum 22
  - Educational and Well-Being Impact of Intrinsic Motivation 23
- Review of Key Points and Hypotheses 24
  - Intrinsic Motivation is Critical for Flow 24
  - Sentiments May be Critical Building Blocks for Internalization 25
  - GIFT Measures Emotions Thoroughly 26
  - Clarification is Needed on the Volatility of Sentiment 26

Chapter 3: Method 28

- Participants 28
- Materials 28
- Design and Procedure 30
  - Process for Study Volunteers 30
  - Measures 30
- Analysis 32

Chapter 4: Results 34

- Pre-task Topic Net Sentiment Impact 35
  - Effect on Internalization 35

Effect on Flow Potential 36
Effect on Information Recall...38
Impact of Pre-task CBI Net Sentiment 38
Effect on Internalization 38
Effect on Flow..40
Effect on Information Recall...41
Impact of Post-Task Sentiments 41
Effect on Internalization 42
Effect on Flow Potential 43
Effect on Information Recall...44
Other Relevant Findings...45
Positive and Negative Sentiment Impact on Information Recall...................................45
Multiple Linear Regression (MLR) Analyses with Post-task Sentiments 46
Topic Sentiment Relationship with Internalization Components 49
Change in Sentiment vs Experience 50
Time on Task Relationships...50

Chapter 5: Discussion and Conclusion 52

Potential Explanations for Variability 52
Wrong Expectation 52
Self-regulation..53
Other Sentiments..53
Extrinsic Incentive 53
Distraction 54
Implications for ITSs 54
Recommendation 1: Use emotion data to better understand learner intrinsic motivation, flow potential, and information recall. 54
Recommendation 2: Select a Sentiment Measurement Strategy that Fits the Tutor's Purpose 55
Recommendation 3: Account for the impact of prior sentiments 55
Recommendation 4: Sentiment shift should be a key metric to monitor to avoid a negative sentiment spiral or to create a positive sentiment spiral 56
Recommendations for Future D2P Development 56
Increase Question Layout Flexibility 57
Add Duplication Capability 57
Enhance Learner Data Export 57
Enable Access Without Sign in Option 58
Areas for Future Research 58
Sentiment and Emotion Sensing Technologies...58
More Experienced Tasks 58
Impact of Other Sentiments 59
Sentiment and Attitudes 59
Concluding Remarks..59

References 61

LIST OF FIGURES

Figure 2.1: Gift Ontology (Sottilare et al., 2012) 7

Figure 2.2: The OCC model of emotion (Ortony et al., 1990) 9

Figure 2.3: Sentiment Formation and Attitude Influence Applied in an OCC Theory Context 13

Figure 2.4: Flow Illustration (Nakamura and Csikszentmihalyi, 2014) 19

Figure 2.5: The Self-Determination Continuum (Ryan and Deci, 2000) 23

Figure 3.1: Camera Shots Tutor Page Example 30

Figure 3.2: Transition to PANAS Questionnaire 31

Figure 3.3: Opening of Modified PANAS Within D2P Tutor 31

Figure 4.1: Total Pre-task Topic Net Sentiment vs Internalization 36

Figure 4.2: Total Pre-task Topic Net Sentiment vs Flow Potential 37

Figure 4.3: Total Pre-task CBI Net Sentiment vs Internalization 39

Figure 4.4: Total Pre-task CBI Net Sentiment vs Flow Potential 41

Figure 4.5: Total Post-task Topic Net Sentiment vs Internalization 43

Figure 4.6: Total Post-task CBI Net Sentiment vs Flow Potential 44

Figure 4.7: Pre-task Negative Sentiment vs Information Recall 46

Figure 4.8: MLR Regression Test to Predict Internalization with Post-Task Sentiments 47

Figure 4.9: MLR Regression Test to Predict Flow Potential with Post-Task Sentiments 48

Figure 4.10: MLR Regression Test to Predict Information Recall with Post-Task Sentiments 49

LIST OF TABLES

Table 4.1: Summary of Correlation and Power Value Results....34

Table 4.2: Descriptive Statistics of Quiz Scores....45

Table 4.3: Correlation Summary of Topic Sentiments vs Internalization Components....50

Table 4.4: Time on Task Correlation Summary....51

ACKNOWLEDGEMENTS

This thesis could not have been accomplished without the support of several mentors and colleagues in the College of IST. First, I'd like to thank Dr. Frank Ritter for serving as my advisor. Dr. Ritter is an advisor who always makes time for his students and encourages cross disciplinary research. Working with Dr. Ritter, I was able to identify and refine a niche research area that I was truly passionate about. His mentorship has made this paper a fun, rewarding experience filled with many life lessons. I am deeply grateful for his mentorship.

Next, I'd like to thank the other members of my committee: Korey MacDougall, Edward Glantz, and Luke Zhang. Korey MacDougall was a mentor for me throughout this entire process. His generous support helped improve the quality of this paper and my overall research experience. Dr. Glantz and Dr. Zhang helped promote my research study and served on my committee. I feel very fortunate to have had their support and feedback.

I also want to express gratitude to Dr. Michael Hills, Sydney Montgomery, Farnaz Tehranchi, and Jake Oury. Dr. Hills generously promoted the research study to his class and offered extra credit. Sydney Montgomery created the written content for the Camera Shots tutor and provided excellent feedback on the tutor's design. Farnaz Tehranchi and Jake Oury reviewed the paper and provided helpful advice throughout the process.

Finally, I'd like to thank my parents. Their love and support is why I have been able to learn and grow at Penn State.

Chapter 1: Introduction

When preparing to take a standardized test such as the GRE, some people are willing to pay hundreds and possibly thousands of dollars for their own private tutor. This is reasonable, as one-on-one tutoring has been shown to be more effective than classroom instruction (Bloom, 1984). A one-on-one tutor can personalize lessons to a student's personal needs and adapt instruction methods to a student's mental state. The effectiveness of one-on-one tutoring has sparked advancement into intelligent tutoring systems (ITSs) research. ITSs are computer-based tutors that provide immediate and customized instruction or feedback to users. ITSs have been shown to be effective in many contexts and have even been implemented in schools (Kulik and Fletcher, 2016).

Modern ITSs seek to replace human tutors (Bloom, 1984; Ma, Adesope, Nesbit, & Liu, 2014), but can an ITS compete with a human teacher in affective areas like emotion sensing, motivating, and engaging? Today, ITSs can identify and adapt to user emotions due to advances in emotion detection software and devices (Arroyo et al., 2009; Woolf, Arroyo, Cooper, Burleson, and Muldner, 2010). These technologies include facial expression sensors, mouse pressure sensors, and posture sensors. Additionally, through applying insights such as attitudes, personality, and goals from user profiles to a theory of emotion formation, tutors are also able to predict potential emotional states in different learning situations (Sottilare, Graesser, Hu, and Holden, 2013).

Despite advances in emotion recognition and adaptation, ITSs have not learned to recognize, predict, and adapt to learner intrinsic motivation and engagement. A human tutor can sense if a student will be intrinsically motivated and engaged in a certain lesson. An ITS cannot do this currently and, therefore, cannot adapt accordingly to foster intrinsic motivation and engagement. Motivation-adaptive learning presents a major opportunity for ITSs to improve the student's learning performance, learning experience, and drive to continue learning the material.

Fortunately, progress in emotion sensing may have unlocked insight into predicting intrinsic motivation and engagement. When discussing emotion-adapting ITS features in her talk *Building an Affective Learning Companion*, Rosalind Picard predicted that emotion-adaptation features "can contagiously excite learners with passion for a topic, leading to greater efforts on the part of the learner to master the topic" (Picard, 2006). Picard's prediction invites investigation into how learner emotions relate to intrinsic motivation and engagement.

In a computer-based instruction context, this research aims to explore how sentiments, which are long-term emotional associations based on prior emotional experiences, relate to intrinsic motivation and engagement theories. I begin by investigating current ITS design recommendations and emotion sensing capabilities. Next, I define sentiment and distinguish its qualities among similar constructs. Finally, I discuss current motivation and engagement research in the context of self-determination theory (Ryan and Deci, 2002) and flow theory (Nakamura and Csikszentmihalyi, 2014). Through a computer-based tutoring experiment, I then test whether learner sentiments have a linear relationship with intrinsic motivation, engagement, and information recall ability.

Chapter 2: Literature Review

This chapter has five sections: Intelligent tutoring systems, sentiment, engagement and flow, intrinsic motivation and internalization, and a review of key points.

Intelligent Tutoring Systems (ITSs)

Intelligent tutoring systems (ITSs) are a type of computer-based instruction (CBI) that are highly adaptive, interactive, and learner-paced as a result of computational models developed in the learning sciences, cognitive sciences, computational linguistics, artificial intelligence, and other relevant fields (Steenbergen-Hu and Cooper, 2014). Recent ITSs have been designed with the intention to replace a human tutor (Ma et al., 2014).

Effectiveness of CBI and ITSs

A meta-analysis conducted by Bernard et al. (2014) revealed that CBI had an effect size of 0.31 as the primary method of learning when compared to a traditional classroom instruction. According to Ma et al. (2014), CBI generally offers key advantages over traditional classroom learning such as:

1. Greater immediacy of feedback (Azevedo and Bernard, 1995)
2. Feedback that is more response specific (Sosa, Berger, Saw, and Mary, 2011)
3. Greater cognitive engagement (Cohen and Dacanay, 1992)
4. More opportunity for practice and feedback (Martin, Klein, and Sullivan, 2007)
5. Increase learner control (Hughes et al., 2013)
6. Individualized task selection (Corbalan, Kester, Jeroen, Van, and Nboer, 2006)

Looking specifically at ITSs, research has suggested that ITSs can outperform classroom learning in certain contexts (Steenbergen-Hu and Cooper, 2014; VanLehn, 2011). Additionally, studies suggest that ITSs can outperform human tutors in more technical learning disciplines (Ma et al., 2014), although, this is not always the case (VanLehn, 2011). Steenbergen-Hu and Cooper (2014) analyzed 39 studies evaluating the use of ITS and college students' academic learning, which resulted in an overall moderate positive effect of 0.35. Based on 27 evaluations, VanLehn found a moderate effective size of 0.76 for step-based tutoring (VanLehn, 2011), and Kulik and Fletcher found a moderate effective size of 0.75 for 39 properly aligned studies (Kulik and Fletcher, 2016). The most effective and widely used ITS has been the Cognitive Tutor, which has shown to increase student test scores by 0.38 standard deviations (Kulik and Fletcher, 2016). Generally, ITSs are praised for three key advantages: immersion, interactivity, and individualization (Fletcher, 2017). Although conflicting results exist (Steenbergen-Hu and Cooper, 2013) and the effect size of ITSs are debated (Kulik and Fletcher, 2016), it is evident that both CBI and ITSs have demonstrated effectiveness in several domains and contexts.

ITS Components

It is not within the scope of this paper to discuss the technical details of how an ITSs is designed. Instead, I will discuss the high-level structure to illustrate what aspects of the tutor this research influences. It is generally accepted that ITSs have four components: The domain model, the learner model, the tutor model, and the user interface model (Graesser, Conley, and Olney, 2012).

The Domain Model

The domain model includes the skill sets, knowledge, and strategies of the topic being taught by the tutor (Sottilare et al., 2013). This component will contain expert-level knowledge of the discipline and common errors learners make while learning. Consider a geometry tutor as an example. The domain model will contain key geometry concepts, formulas, and frequent errors exhibited by learners.

The Learner Model

The learner model contains the cognitive, affective, motivational, and other mental states that the learner will experience during learning (Sottilare et al., 2013). The learner's progress is also stored in this component along with learner strengths and weaknesses. Continuing with the hypothetical geometry tutor example, the learner model would house insights like the learner's current skill level with geometry, attitude toward geometry, emotional states, and weaknesses. This research aims to influence this component; thus, I will discuss it in further detail in the following section.

The Tutor Model

The tutor model combines insights from the learner model and the domain model to determine the ideal instructional methods (Sottilare et al., 2013). These methods include instructional steps, strategies, and actions based on the evolving domain and learner model. In the hypothetical geometry example, consider a scenario where the learner is struggling with area formulas due to a weakness in basic mathematical operations. Referencing the learner's profile, the tutor knows that the learner has an affinity for video games. As a result, the tutor model could adapt and direct the learner to a beginner-level mathematical operations video game. After the game, the tutor will assess the state and determine the next optimal instructional strategy given the new inputs.

The User Interface Model

The user interface model adapts to the learner contributions through a variety of input types to determine the ideal way to improve the user experience (Sottilare et al., 2013). These inputs could include mouse movements, mouse clicks, typing, speech, and many others. Note that the difference between the user interface model and the tutor model is that the user interface models is focused on the usability for the learner. Continuing the hypothetical geometry tutor example, the tutor might determine through eye tracking that the user is taking a long time to find the problem on the page. As a result, the tutor could adapt to show only one problem per page and change the font size and color of the question text.

The Learner Model and ITS Standardization

As this research is investigating the impact of user sentiments on key learning outcomes, the focus will be on the learner model (also often called the student model). Learner model design has varied among ITSs. Early ITS systems were built independently, each with different goals and designs (Sottilare, Brawner, Goldberg, and Holden, 2012). These early systems were difficult to adapt and had limited reusability (Picard, 2006). This approach proved to be expensive and impractical for organizations to develop and implement within a reasonable budget (Picard, 2006). Thus, ITS researchers have worked to create standards for ITSs development. I will be focusing on the learner model design recommendations of one of those standards called the

Generalized Intelligent Framework for Tutoring (GIFT). Although this research provides insight relevant to other architectures such as Cognitive Tutor Authoring Tools (CTAT) or Declarative to Procedural (D2P) (Ritter et al., 2013), I will focus on GIFT given its focus on affective states in the learner model.

Generalized Intelligent Framework for Tutoring (GIFT)

Created in 2012, GIFT is an empirically-based, service-oriented framework of resources and standards that facilitates authoring, managing, and assessing elements of an ITSs (Sottilare et al., 2012). Although the GIFT initiative is military led, it is being designed for use in industry and academia as well as military training (Sottilare et al., 2013).

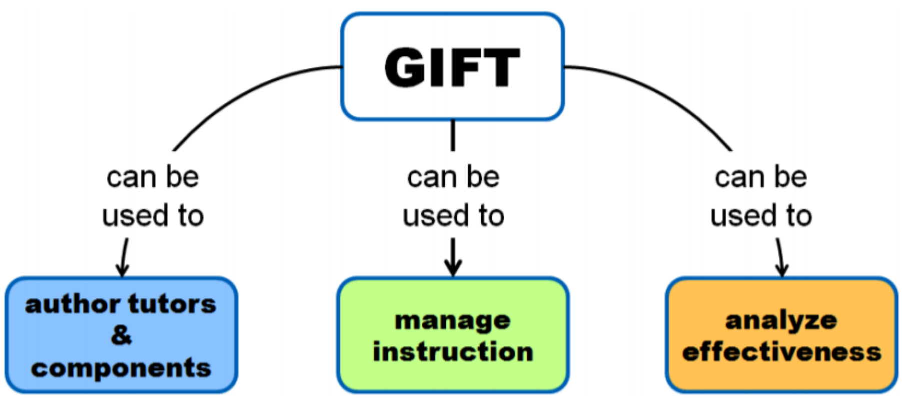

Figure 2.1: Gift Ontology (Sottilare et al., 2012)

The ABC User Model

Within GIFT, the learner model uses an Affective-Behavioral-Cognitive (ABC) User Model to measure and predict affective, behavioral, and cognitive states and patterns (Sottilare et al., 2013). Sottilare et al. (2013) gives an overview of the purpose and inputs of each component.

The affective component stores information relevant to user emotions. I will discuss this component in greater detail in the next section. The behavioral component stores information about how the user interacts and reacts to events. Inputs of the behavioral component include personality traits based on the Five-Factor Model of Personality, different types of goals, and levels of goal intensity. The cognitive component stores information about the user's mental process. Inputs of the cognitive component include cognitive state thresholds, cognitive state decays, predicted cognitive states, and active cognitive states.

Affective Component of the ABC User Model

Ortony, Clore, and Collins Cognitive Appraisal Theory

Although there is debate in how emotions are formed (Panksepp, 2003), The ABC User model requires a cognitive theory of emotions to serve as its affective foundation. It uses Ortony, Clore, and Collins' cognitive appraisal theory (OCC Theory) to serve as the basis of the affective aspect of the ABC User model (Sottilare et al., 2013).

The OCC theory is one of many appraisal theories. The core of appraisal theories is that emotions are elicited according to an individual's subjective interpretation or evaluation of important events or situations (Roseman and Smith, 2001). A key theme of appraisal theory is that a cognitive evaluation (the appraisal) comes first and therefore creates the emotion. This appraisal process occurs multiple times throughout the duration of a task. Thus, appraisal theory is often applied in self-regulation research as the high frequency of situation appraisals are critical to understanding how emotional change occurs (Gross, 2015).

In the OCC theory, an appraisal structure has three components that determine the emotion and the emotion's strength. These components are the individual's goals, attitudes, and standards (Ortony, Clore, and Collins, 1990). The intensity of emotions stemming from goals is desirability; the intensity of emotions stemming from attitudes is appealingness; and the intensity of emotions stemming from standards is praiseworthiness (Ortony et al., 1990). According to the OCC theory, people not only appraise objects but also events, cognitive states, and emotional states (Ortony et al., 1990). It may sound strange that individuals appraise emotional states, but this can be seen often in everyday life. Consider the person who is abnormally afraid of fear itself or the person who becomes overly happy when they are experiencing love.

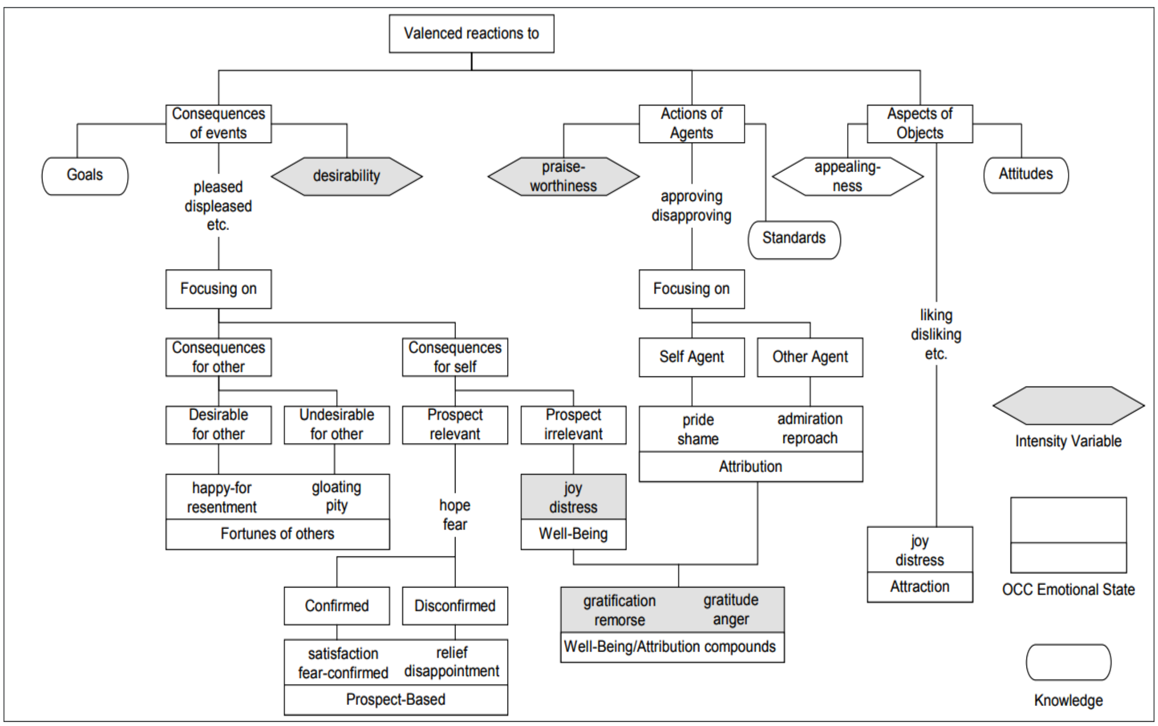

Figure 2.2: The OCC model of emotion (Ortony et al., 1990)

When faced with a complex task such as learning Shogi (Japanese chess) on a computer in a dorm room with rock music in the background, it becomes apparent that multiple appraisals could be occurring simultaneously. The OCC theory claims that an emotion must be *valenced* (Ortony et al., 1990). Valenced means it must have a positive or negative charge. Excitement, for example, has positive valence while fear has negative valence. Thus, the *overall task emotion* *valence* is the combination of the valences of all emotion appraisals that occurred during a task of interest.

Emotional Profile

The emotional profile is a collection of inputs within the affective component. It is responsible for representing emotional patterns demonstrated by the learner (Sottilare et al., 2013). It includes the inputs of emotional class thresholds and emotional class decays. Emotional class thresholds are predicted emotional resistances the user has for different emotion classes. Sottilare et al. (2013) use the example of measuring how disappointed a learner can feel upon not solving a problem. The other section of the emotional profile is called the emotional class decays section. It collects data on how long certain emotions last for the learner.

Emotional States

The emotional states section of the affective component collects data on two types of emotional states. The first is the potential emotional states. This section uses the OCC theory and user information to predict how the learner will feel in a particular learning situation (Sottilare et al., 2013). Second is the active emotional states. This is determined using emotion sensing technology (Arroyo et al., 2009; Woolf et al., 2009) to determine how the learner is feeling in real time. The potential state is compared to the active state to improve future potential state predictions (Sottilare et al., 2013).

Sentiment

Research into emotion and learning suggests that positive emotions lead to both intrinsically motivated and engaged behavior (Schutz and Pekrun, 2007). Thus, this research seeks to better understand what motivation and engagement insights can be gained through analyzing emotion data stored in the affective component of the leaner model's ABC User model. As there are numerous emotion appraisals during a learning task, this research needs a broader construct than emotion to predict intrinsic motivation. However, it cannot be too broad and incalculable with the recommended ABC User Model guidelines. As a result, this research focuses on the impact and nature of sentiments.

Sentiment Overview

Sentiments are long-term emotional associations toward an object that are a result of the valence of previous and expected emotional experiences with the object (Brave and Nass, 2009). Sentiments, like appraised emotions, must be valenced. Note that the term sentiment is not consistent across academic disciplines. Depending on the research discipline, sentiment research is conducted using related constructs like emotion, mood, attitude, and interest. This research choses to investigate sentiment for two reasons. First, given the design of the ABC model in ITSs, I predict that a sentiment measurement can be produced through manipulating the emotional state data of the ABC User Model. Second, as sentiments are more long-term and resistant to change, I predict they will be more predictive of intrinsic motivation, engagement, and even information recall.

As GIFT recommends OCC theory for its emotion formation theory, it is important to understand how sentiments are formed in the context of OCC theory. In the Ortony, Clore, and Collins Cognitive Appraisal Theory section, I outlined that the overall task emotion valence is the combination of the valences of all emotion appraisals that occurred during the duration of the task. These overall task emotion valences are the building blocks for sentiments. I define sentiment for a task as the combination of all overall task emotion valences that have been experienced in a person's lifetime for that task.

For example, consider a person who is playing with a cute, friendly puppy. While he is playing with the puppy, he is most likely experiencing frequent positive emotion appraisals. As a result, his overall task emotion valence for the task of playing with that puppy that one time will be positive. If throughout his life, this person consistently has positive overall task emotion valence for playing with puppies, he will have a positive sentiment for playing with puppies. Valenced emotion appraisals build overall task emotion valences, which, in turn, build sentiments (Figure 2.3). Note that with no task experience, I assume individuals will use an expected overall task emotion valence to determine their sentiment. For instance, if someone has never played Shogi (Japanese Chess), he or she would likely apply his or her sentiment of American chess to determine the expected sentiment for Shogi.

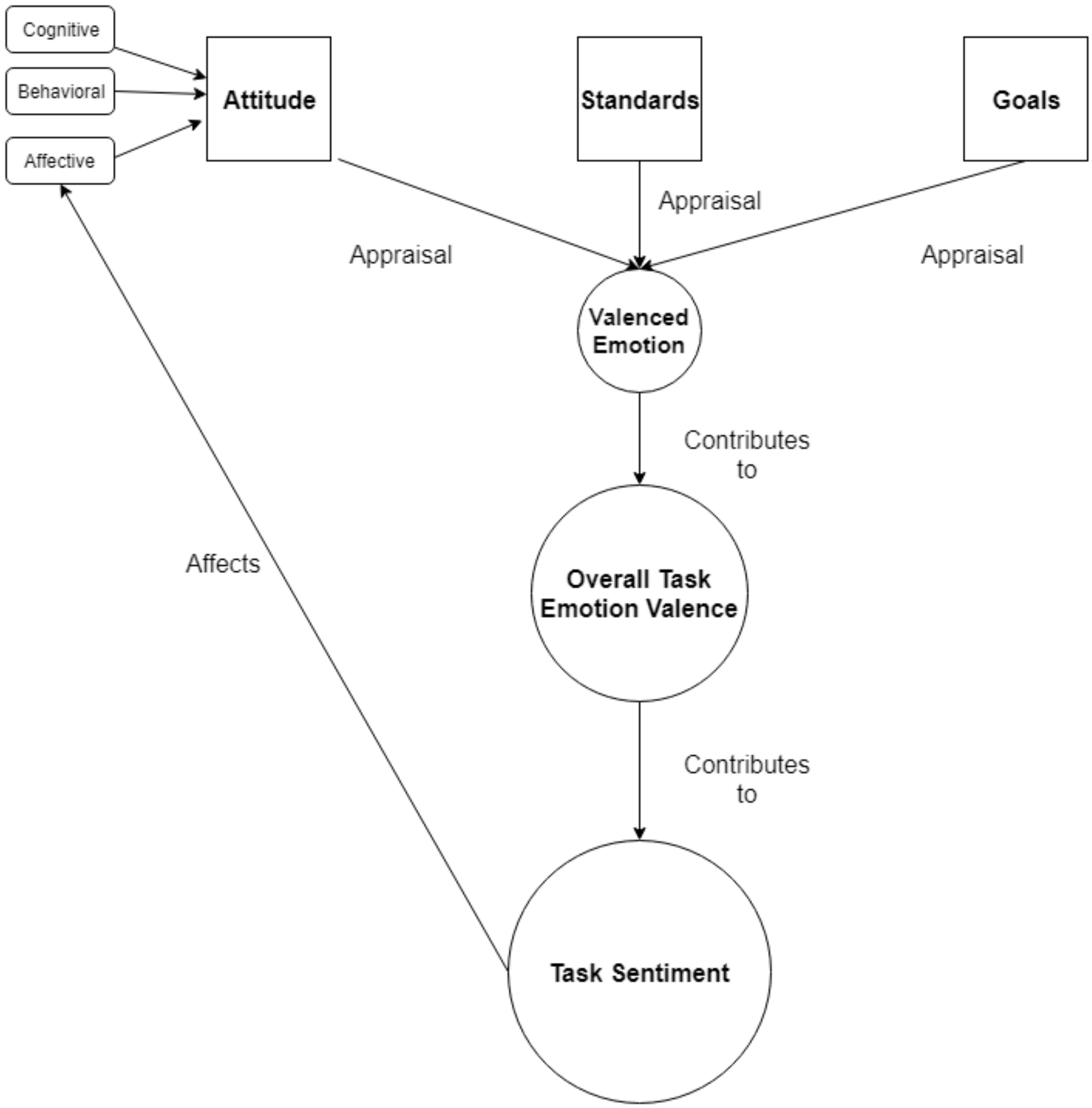

Figure 2.3: Sentiment Formation and Attitude Influence Applied in an OCC Theory Context. In this figure, I added Overall Task Emotion Valence and Task Sentiment concepts to a simplified version of Figure 2.2 and to a visualization of the components of attitude (Breckler, 1984).

Distinguishing Between Related Terms

The usage of the term sentiment and other related terms vary across research disciplines. As sentiment was described in the last section, this section defines and explains terms similar to sentiment.

Affect

Affect is a broad category that encompasses emotion-related constructs such as emotion, mood, sentiment, emotion-regulation, and interest (Woolf et al., 2009). However, in emotion and learning research, it is very common to see the terms “positive affect” and “negative affect.” These terms typically refer to short-term emotion and mood valences as oppose to long-term sentiment or attitude valences (Mageau & Vallerand, 2007; Ryan, Bernstein, & Warren Brown, 2010; Watson, Clark, & Tellegen, 1988). It is noteworthy that the definition of affect seems to shift by not only research fields but by researchers. For this research, affect is a broad category that encompasses multiple constructs.

Emotion

Brave and Nass (2003) provides an overview of the differences between emotion and sentiment in the context of HCI. Similar to sentiments, emotions are directed toward a particular object, for instance, a person is mad at someone, nervous about an event, or loves something. Unlike sentiments however, emotions are fleeting in nature. While sentiments are long-term associations, emotions are short term feelings that disappear in minutes or hours. For example, a person could experience frustration while playing a video game, but he or she could still have a very high sentiment for video games. Despite their fleeting nature, emotions, according to Brave and Nass (2003), play a key role in developing sentiments. To use the same example, a person who typically experiences frustration while playing video games would most likely have a reinforced negative sentiment for video games. The ABC User model has a framework and guidelines to measure emotions, and because emotions are the building blocks of sentiments, I predict a sentiment calculation is within the capability of GIFT's current ABC User model.

Emotion appears to be the most common affective construct that is compared to motivation and engagement. Research in multiple disciplines suggests that emotion affects motivation, engagement, and learning (Bradley, 2000; Erez and Isen, 2002; Graham, 1991; Mageau and Vallerand, 2007; Meyer and Turner, 2006). Bradley (2002) discussed effects of emotions from a neuroscience perspective and highlights how emotions especially impacts attention. Graham (1991) compares emotions to attribution theory and concludes "A viable theory of motivation for educational psychology must be able to incorporate emotions." In teaching strategies, Turner, Meyer, Midgley, and Patrick (2003) emphasize the importance of a positive teaching style on learning outcomes, including motivation and engagement. This can be seen an ITS research as well with affective learning companion research (Picard, 2006).

Mood

Moods are slightly more long-term than emotions but are more temporary than sentiments and do not have an object (Brave and Nass, 2009). For instance, *Bill is happy* is an example of a mood. He is not happy about some object, he is just happy. Mood is often referred to as an emotional state and has shown to have motivational effects (Kavanagh and Bower, 1985).

Attitude

Attitude is a broader concept than sentiment. It is generally accepted that attitudes have three components: the cognitive component, the affective component, and the behavioral component (Ajzen and Fishbein, 1977; Breckler, 1984). An individual's sentiment would influence the affective component of an individual's attitude. In fact, sentiment is nearly identical in nature to the affective component of attitude. This is sometimes called affective attitude (Connor, Sheeran, Godin, and Germain, 2013). Thus, I assume change in sentiment will shift an overall attitude (Figure 2.3). This is supported by Edwards (1990) claim that affective appeals are more likely than cognitive appeals to shift overall attitude. This is relevant because, as mentioned prior, attitude is one of three components of emotion appraisal in OCC theory.

Interest

Similar to attitude, interest is a broader concept than sentiment. In general, researchers have separated interest into two types: individual interest and situational interest (Hidi, 1990). Situational interest is an increased interest due to environmental factors. An example of leveraging situational interest is a science teacher who lights something on fire to get his or her class's attention. Individual interests are more stable and are based on individual evaluations of objects (Hidi, 1990). Two components of individual interest have been distinguished by researchers: feeling-related valence and value-related valence.

Feeling-related valence

The feeling-related valence aspect of interest theory is a similar concept to sentiment. This type of valence refers to the feelings that are associated with an object or an activity (Wigfield, Eccles, Schiefele, Roeser, and Davis-Kean, 2008). Note that feeling-related valence is only one aspect of overall interest, which is why interest is considered a broader construct than sentiment. Like sentiment, there is a clear connection between feeling-related valence of interest theory and the affective component of attitude.

Value-related valence

Value-related valence refers to the attribution of personal significance of the object (Wigfield et al., 2008). For instance, an environmentalist may have a high value-related valence toward learning about solar panels because there is personal significance. Value-related valence relates to the cognitive component of attitude as opposed to the affective.

Engagement and Flow

One objective of this research is to explore if sentiments can predict task engagement. There are a multitude of different engagement theories across different academic disciplines. Each engagement theory can typically be classified as one or more of the following: cognitive engagement, emotional engagement, or behavioral engagement (Fredricks, Blumenfeld, and Paris, 2004). This research utilizes flow theory as its basis for measuring engagement.

This research picked flow theory for three reasons. First, flow is considered a state of complete engagement (Csikszentmihalyi, 1990), which should be the level of engagement that ITSs should be striving to produce. Second, flow encompassed all three types of engagement (cognitive, emotional, and behavioral). Finally, flow is a construct that is commonly seen in CBI research (Challco, Andrade, Borges, Bittencourt, & Isotani, 2016; Choi, Kim, & Kim, 2007) due to a computer's ability to adapt to the learner.

Flow Theory Overview

Flow theory has been applied to a broad range of contexts, including education, sports, work, shopping, rock climbing, dancing, and games (Csikszentmihalyi and Lefevre, 1989). Originally, the concept of flow was defined as the holistic sensation people experience when they act with complete involvement (Csikszentmihalyi and Lefevre, 1989). In other words, flow is the ultimate form of absorption and engagement in an activity. Nakamura and Csikszentmihalyi (2014) states that being in the flow state includes the following characteristics:

1. Intense and focused concentration on what one is doing in the present
2. Merging of action and awareness
3. Loss of reflective self-consciousness
4. A sense that one can control one's actions
5. Distortion of temporal experience, with the typical case being time went faster than normal
6. Experience of the activity as intrinsically rewarding

The blend of these psychological states creates a deep sense of enjoyment that people strive to reproduce (Csikszentmihalyi, 1990). The drive toward reproducing flow can even result in addictive behaviors. This has been especially prevalent in gaming (Wan and Chiou, 2006).

There are several key conditions to enter the flow state. First, the activity must be an optimal balance of skill and difficulty (Nakamura and Csikszentmihalyi, 2014). Flow cannot occur if a task is too hard or if the task is too easy. Second, the user must have clear proximal

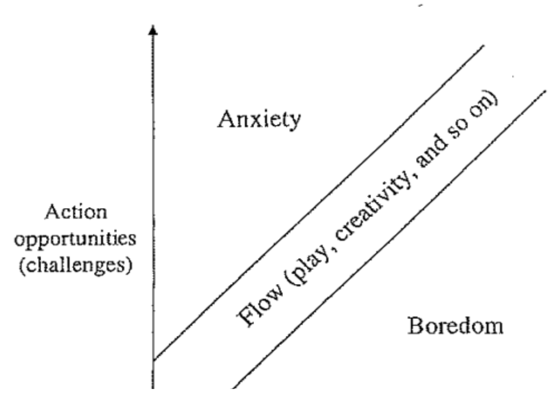

Figure 2.4: Flow Illustration (Nakamura and Csikszentmihalyi, 2014)

goals and the task gives immediate feedback on the fulfillment of those goals (Nakamura and Csikszentmihalyi, 2014). Finally, the individual must be intrinsically motivated to perform the task enter flow (Keller and Bless, 2008). This intrinsic motivation condition is often broken into three sub-conditions: concentration, interest, and enjoyment (Shernoff and Csikszentmihalyi, 2009).

Flow and Education

Due to numerous learning advantages, engagement in both classroom and online settings has been an area of focus in the field of education (Shernoff and Csikszentmihalyi, 2009). As the optimal form of engagement with a tendency to produce a strong euphoria, flow has also been a focus. For instance, the Key School in Indianapolis is a school devoted to creating flow by influencing both the environment and the individual (Nakamura and Csikszentmihalyi, 2014). Additionally, flow has been considered a key learning outcome in CBI contexts, including in ITSs (Woolf et al., 2009, 2010)

Intrinsic Motivation and Internalization

As flow is contingent on the development of intrinsic motivation, it is critical to understand the current state of motivation research. The nature of motivation has been approached from multiple psychology perspectives including cognitive, developmental, social, and educational (Lazowski and Hulleman, 2015). As a result, an extensive list of constructs and theoretical frameworks have been developed. In general however, motivation theories are concerned with the energization and direction of behavior (Lazowski and Hulleman, 2015). Motivation theories typically answer one or more of the following questions (Wigfield et al., 2008):

1) Can I do this task?
2) Do I want to do this task and why?
3) What do I have to do to succeed on this task?

To adequately assess the impact of sentiment on intrinsic motivation in the CBI context, this research required a single motivation theory as a foundation. I selected Self-Determination Theory (SDT) for three reasons. First, there is significant application of SDT in the learning context, including CBI. For instance, Reeve and colleagues successfully integrated a SDT-based intrinsic motivation intervention within a computer-based conversational Chinese tutor (Reeve, Jang, Hardre, and Omura, 2002). The results illustrated that the intervention increased intrinsic motivation to learn conversational Chinese. Second, I sought a motivation theory that addressed all three questions typically addressed by motivation theories. SDT succeeds in addressing each question through integrating insights from five sub-theories. Finally, because flow relies on the development of intrinsic motivation, I wanted a theory that has a focus on not just increasing motivation but increasing specifically intrinsic motivation. SDT proposes the process of internalization, which represents the active assimilation of behavioral regulations that are originally foreign in nature (Ryan, 1995). In other words, through internalization, a person finds more internal significance or connection in a task, thus increasing intrinsic motivation.

Self-Determination Theory Overview

Self-Determination Theory (SDT) is an organismic theory of development with a focus on the role of motivation in development and learning (Ryan and Deci, 2002). It is a combination of five mini theories: Cognitive Evaluation Theory, Organismic Integration Theory, Causality Orientations Theory, Basic Psychological Needs Theory, and Goals Contents Theory. SDT principles have been applied in a variety of disciplines including education, organizational behavior, human computer interaction, and athletics. Although there are many aspects of SDT, I will focus here on the three basic needs and the self-determination continuum.

Deci, Ryan, and their colleagues claim there are three basic psychological needs that contribute to intrinsic motivation: autonomy, competence, and relatedness (Ryan and Deci, 2002). Autonomy is related to the source of one's own behavior. This need refers to acting as a result of interest and intrinsic values as opposed to instrumental reasons. Competence refers to feeling effective in social interactions and experiences in order to express one's ability. This need is similar to Bandura's Self-Efficacy Theory, which claims that one's belief of effectiveness toward a particular event will affect their drive to participant in that event (Bandura, 1977). Finally, relatedness refers to feeling connected to other people and includes caring for others and being cared for by others. SDT claims that a cognitive evaluation of how much a task will satisfy each of these three needs will contribute to determining intrinsic motivation (Ryan and Deci, 2000). Moreover, improving these needs will contribute toward internalization (Ryan, 1995). In education, this claim has been supported by several studies showing successful motivation interventions that target an increase in competence, autonomy, or relatedness (Lazowski and Hulleman, 2015).

The Self-Determination Continuum

SDT proposes four types of extrinsic motivation that vary in terms of autonomy. These are outlined in Figure 2.5. Before extrinsic motivation however, there is amotivation. Amotivation occurs when there is neither extrinsic nor intrinsic incentives for a person to engage in a task. The locus of causality is impersonal become there is no personal incentive for the individual. External regulation (low autonomy) involves external rewards or punishments (e.g., “I’ll give you $50 to do this”). Introjected regulation (moderately low autonomy) involves pressure from an individual’s ego (e.g., “All the cool kids are on the football team, so I will join the football team”). Identified regulation (moderately high autonomy) involves finding relevance or importance in an activity (e.g., “I will learn statistics because it is critical for what I want to do in life”). Finally, integrated regulation (high autonomy) involves an activity that synthesizes with intrinsic motivations (e.g., “I love being a politician and sharing my beliefs, so I will write an election speech even though I do not like writing”). Moving along this continuum is internalizing an action. It is noteworthy that internalization does not need to follow this continuum as the process of internalization depends on past experiences (Ryan, 1995). For instance, a person could move from external regulation to identified regulation and skip introjected regulation.

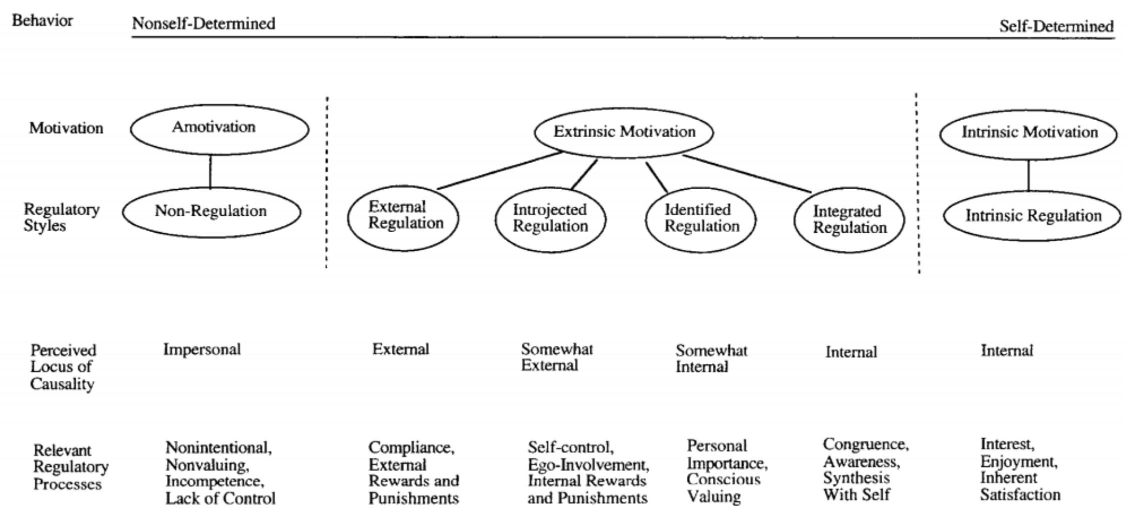

Figure 2.5: The Self-Determination Continuum (Ryan and Deci, 2000)

Educational and Well-Being Impact of Intrinsic Motivation

There have been multiple studies supporting that SDT interventions and principles can support both intrinsic motivation to learn and learning performance (Niemiec and Ryan, 2009). For instance, Benware and Deci (1984) had college students learn science material. One group had to learn in order to teach other students while the other group had to learn to pass a test. The group that learned to teach was shown to had experienced greater conceptual learning and intrinsic motivation to learn the material (Benware and Deci, 1984). Focusing on the three core needs presented by SDT has also shown to increase positive emotions and overall wellbeing (Ryan and Deci, 2000). For instance, a study by Ryan, Bernstein, and Brown measured affect in working-class adults throughout the week. The study also measured throughout the week when the needs of autonomy, relatedness, and competence were being met. The results showed that points in the week of high positive affect matched with periods where needs were more significantly met (Ryan, Bernstein, and Warren Brown, 2010). This is especially relevant to learning given research that shows positive emotions while learning provides multiple learning advantages, including enhanced information recall (Schutz and Pekrun, 2007). The learning benefits of SDT have been shown to apply across different cultures as well (Niemiec and Ryan, 2009). Tsai et al. (2008) illustrated that seventh grade German students experienced higher levels of interest in three subjects when they had autonomy-supportive teachers. Similarly, Jang, Reeve, Ryan, and Kim (2009) showed that South Korean public school students experienced higher intrinsic motivation for learning upon feelings of autonomy and competence.

Review of Key Points and Hypotheses

The review blended research from ITSs, affect and learning, and motivation and engagement. In this section, I summarize key points of the review and rationalize the hypotheses. There are five key points I want to highlight: Intrinsic motivation is critical for flow, sentiments may be critical building blocks for internalization, GIFT measures emotions thoroughly, and clarification is needed on sentiment volatility.

Intrinsic Motivation is Critical for Flow

Highlighted in Flow Theory Overview, there are three key conditions to enter Flow: Balance of skill and difficulty, goals and immediate feedback on goal progress, and intrinsic motivation for the task. An ITS can account for the first two conditions. For instance, Anderson and colleagues provided eight principles for cognitive tutor design that address the first two conditions of flow (Anderson, Corbett, Koedinger, and Pelletier, 1995). Principle one relates to monitoring the user skill level to obtain an ideal balance of difficulty and skill. Principle two relates to goal setting while principle six directly states to provide immediate feedback on errors.

The obstacle for ITSs to enable flow is not the first two conditions, it is the third condition of intrinsic motivation. Because intrinsic motivation is critical to flow and provides several other benefits as discussed in the Educational and Well-Being Impact of Intrinsic Motivation section, it is a major opportunity for ITSs to acquire additional insight into the learner's intrinsic motivation.

Sentiments May be Critical Building Blocks for Internalization

Positive emotions and well-being are documented outcomes of having intrinsic motivation for a task. This includes SDT as discussed in Educational and Well-Being Impact of Intrinsic Motivation section and Flow Theory as discussed in Flow Theory Overview section. As discussed in the Emotion section, multiple studies suggest that positive emotions are predictive of intrinsic motivation and engagement. Thus, there is confusion around which comes first: positive emotions or intrinsic motivation? Similar to how emotions are building blocks for sentiments, I hypothesize that sentiments are building blocks for intrinsic motivation and internalization. Thus, the first hypothesis is:

Hypothesis 1: Topic pre-task sentiment is a positive linear predictor of degree of internalization

Additionally, as intrinsic motivation is a condition for flow, I hypothesize that sentiment will also be predictive of flow potential. This leads to the second hypothesis:

Hypothesis 2: Topic pre-task sentiment is a positive linear predictor of flow potential

It is noteworthy that sentiment for a learning topic is only one sentiment and, thus, is only one hypothesized building block for intrinsic motivation. Therefore, other sentiments surrounding the task are hypothesized to be relevant to the development of intrinsic motivation as well. This research predicts that one of those relevant sentiments is the sentiment toward the medium of learning. The medium here is CBI, so CBI sentiment is hypothesized to be relevant to predicting both the degree of internalization and flow potential. This leads to the third and fourth hypothesis:

Hypothesis 3: CBI pre-task sentiment is a positive linear predictor of degree of internalization

Hypothesis 4: CBI pre-task sentiment is a positive linear predictor of flow potential

GIFT Measures Emotions Thoroughly

As discussed in Affective Component of the ABC User Model section, GIFT provides a framework via the ABC User Model that is capable of emotion measurement. The ABC User Model is able to develop a predicted emotional state through applying OCC theory to their dynamic user profile. In addition, through emotion sensors such as Affectiva, the ABC User Model is also able to measure the actual emotional state. Given emotions are building blocks of sentiments, there is an opportunity for the ABC User Model to use these emotions to calculate the sentiment of the learner. If the research predictions are true, the ABC User Model will then also be able use the calculated sentiment to predict the degree of internalization and flow potential for the user for each learning task.

Clarification is Needed on the Volatility of Sentiment

As emotions are building blocks for sentiments, I hypothesize that how much experience a user has with an object will serve as a predictor of the volatility of the sentiment. This is because more experience entails more emotions and, therefore, more reinforcement. This leads to the final hypothesis:

Hypothesis 5: Experience is a negative linear predictor of sentiment volatility

If the research predictions are true, it is relevant for ITS designers to better understand how sentiments change. For instance, if a critical sentiment is developing and is in a volatile state, ITSs should take action to ensure that positive emotions are experienced. Failure to do this early on could result in early negative sentiment and therefore early amotivation. As negative sentiment will decrease attitude as well, future emotion appraisals surrounding this task will more likely to be negatively valenced.

Chapter 3: Method

Participants

The experiment had a total sample size of 70 volunteers. Although a total of 75 volunteers participate in the study, five provided incomplete data and were removed from analysis. 26 were females, 43 were males, and one did not identify as either male or female. All students were undergraduates from the College of Information Sciences of Sciences and Technology (IST) at The Pennsylvania State University. As an incentive, volunteers who participated received extra credit in one of their IST courses.

Materials

The entire research study was on a D2P tutor (Ritter et al., 2013) that contained lessons and questionnaires. There was a total of 205 questions. Volunteers could access the tutor from any computer or location using a link. The tutor contained two lessons: Introduction to how to play Shogi (Japanese chess) and an introduction to Camera Shots in cinematography. To control for user experience, both lessons had nearly identical visual layouts and usability features.

I selected the lesson topics of Camera Shots and Shogi for three reasons. First, most likely, learning these topics will not be relevant toward any career goal, which prevents intrinsic motivation scores to be influenced by achievement aspirations (cognitive component related). Second, I predicted that students would not have much experience in either Camera Shots or Shogi. This was validated via a self-report measure asking experience level. An even experience level provided a basis for assessing sentiment volatility. Finally, I felt that each lesson appealed to a different personality. Shogi is more technical and tactile while Camera Shots is more visual and artsy.

Both lessons were approximately the same length, but this varied by participant. The Shogi lesson contained the following sections:

1. Introduction to Shogi
2. Pieces Overview
3. Piece Promotion
4. Piece Placement

The Camera Shots lesson contained the following sections:

1. Camera Shots Introduction
2. Camera Shots Key Terms
3. Types of Shots and When to Use Them
4. Basic Rules of Compositions

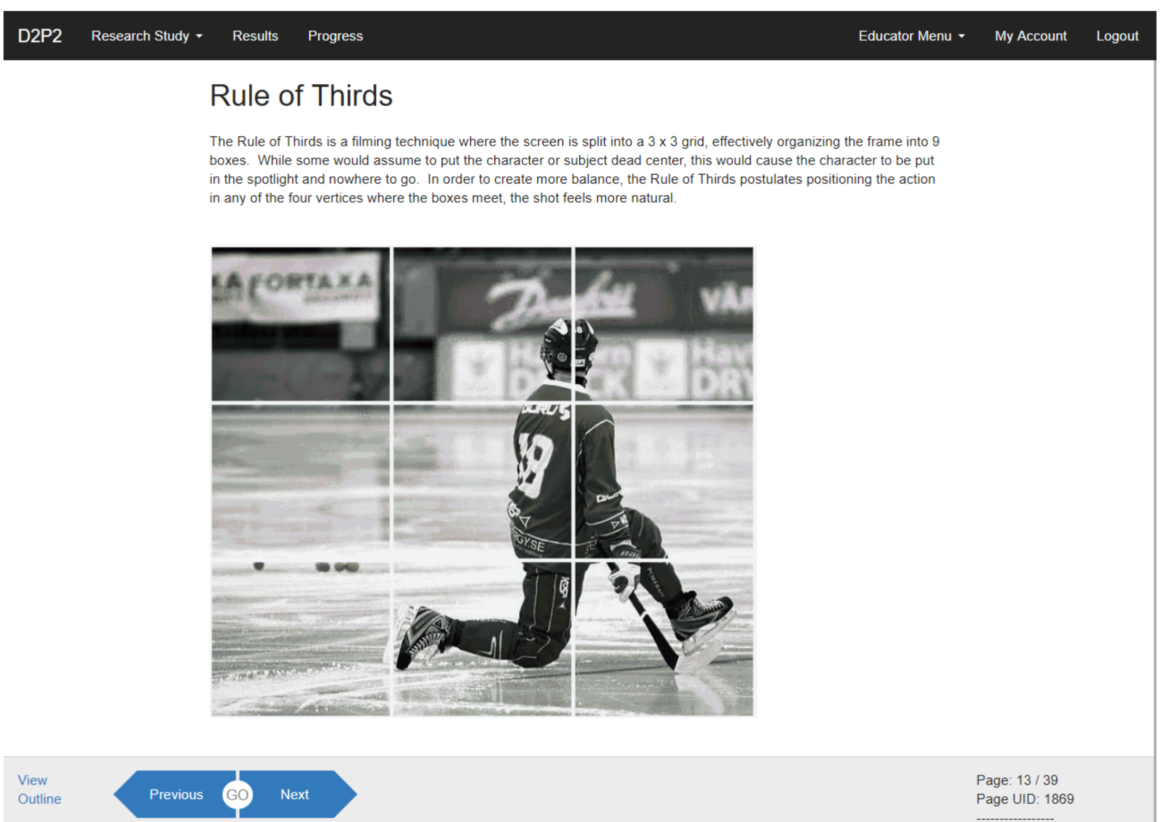

Figure 3.1: Camera Shots Tutor Page Example

Design and Procedure

Process for Study Volunteers

Volunteers who signed up to participate received an email with a link to the study within 48 hours. Volunteers clicked the link to open the study. The study, including the questionnaires and quizzes, was all within a computer-based tutor. Volunteers had three to seven days to complete the study after receiving the link to it and could complete the study at any physical location they desired on any computer. The study took approximately 30 to 60 minutes. There was no time limit. The order of the lessons was switched for 39 of the 75 study volunteers.

Measures

The study began with a pre-test that determined mood, experience levels, college year, CBI sentiment, and gender. After completing the pretest, volunteers could begin the first lesson. Before taking each lesson, volunteers completed a slightly-modified version of The Positive and Negative Affect Schedule (PANAS) to assess their sentiment for the topic (Crawford and Henry, 2004). This is a questionnaire that consists of two 10-item scales and measures both positive and negative emotion; however, the wording was altered to be predictive of sentiment as opposed to emotion. My modified questionnaire asked participations what emotions they felt when they envisioned themselves performing a task as opposed to how they felt at that moment. This was done through a transition paragraph prior to the PANAS page (Figure 3.2).

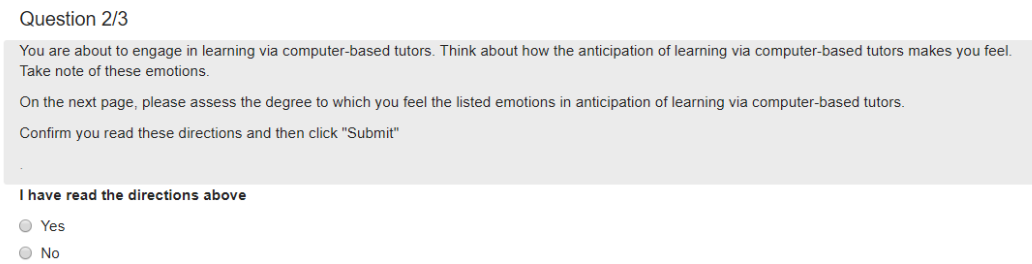

Figure 3.2: Transition to PANAS Questionnaire

The strength of the emotions was then rated from 1 (low strength) to 7 (high strength). Half of the rated emotions were negatively valenced while the other half were positively valenced. After completing both lessons, volunteers completed The Positive and Negative Affect Schedule again for each topic to assess their post-task sentiment and change in sentiment. I selected PANAS because of its popularity in affective research (Brave & Nass, 2009; Mageau & Vallerand, 2007; Ryan et al., 2010). It also measures similar emotions that emotion sensing technology seek to measure (Woolf et al., 2010).

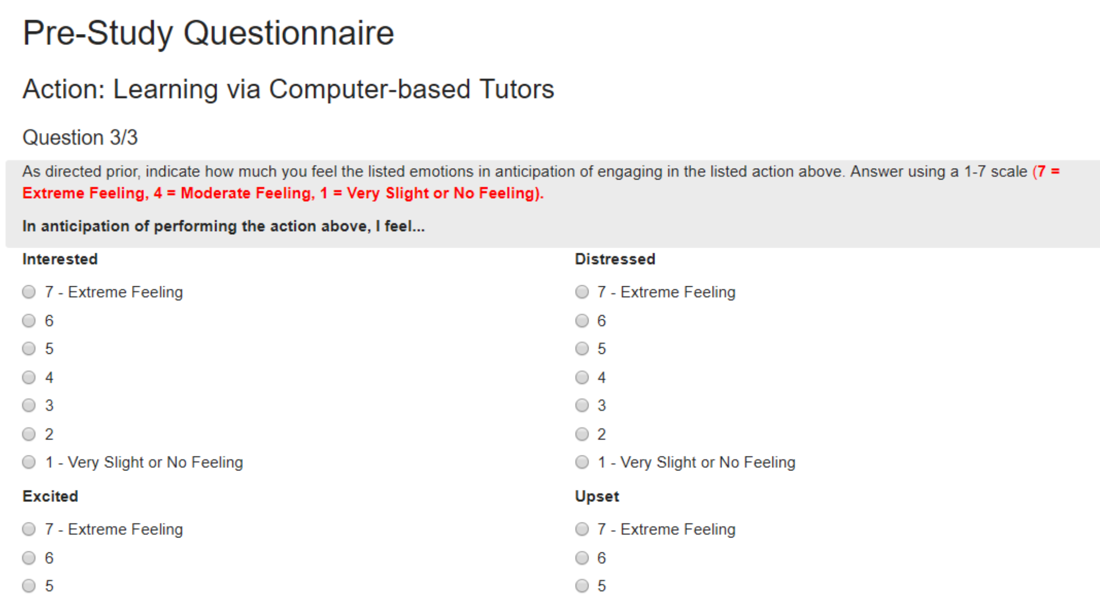

Figure 3.3: Opening of Modified PANAS Within D2P Tutor

After each lesson, volunteers completed an information recall multiple choice quiz, a flow questionnaire, and an internalization questionnaire. Both the Camera Shots and Shogi quizzes were 10 questions that were either multiple choice or true or false. Internalization was measured by assessing and combining perceived topic importance, perceived self-determination, effort, interest, and enjoyment scores. This measurement strategy is based on Self-determination theory and was adapted from Reeve et al. (2002) Chinese conversational tutor study. Engagement was measured using the 10-question Flow Short Scale questionnaire (Jackson, Martin, and Eklund, 2008) minus the tenth question due to its lack of applicability to CBI. This paper refers to the flow measurement results as a flow potential score to avoid the risk of implying that volunteers entered a state of flow. I assume most volunteers did not enter a flow state. Rather, the experiment is assessing how much they demonstrated qualities of a flow state.

Analysis

For both the Camera Shots and Shogi tutors, each participant had eight scores that were calculated using their participation data: Pre-Task Net Sentiment for Topic, Pre-Task Net Sentiment for Method (CBI), Post-Task Net Sentiment for Topic, Post-Task Net Sentiment for Method, Flow Potential, Intrinsic Motivation, Information Recall, and Change in Sentiment. Although Post-Task Sentiments and Quiz Scores were not in my hypotheses, I felt these calculations could potentially provide unexpected insights.

All Task Net Sentiments were calculated by subtracting the sum of negative sentiment from the sum of positive sentiments. Thus, this score could be negative. Note that the Pre-Task Net Sentiment only deals with PANAS data that was collected prior to the participant taking the tutor or lesson while Post-Task Net Sentiment deals with data from PANAS at the end of the study. The Change in Sentiment was calculated by subtracting the Post-Task Net Sentiment from the Pre-Task Net Sentiment.

The Flow Potential score was calculated by summing the scores from the Flow Questionnaire. The Intrinsic Motivation score was calculated by summing the averages of the interest questionnaire, the perceived self-determination questionnaire, the perceived topic importance questionnaire, the effort questionnaire, the interest questionnaire, and the enjoyment questionnaire for each lesson. The impact of Pre-Task Net Sentiment, Post-Task Net Sentiment, and Change in Sentiment on each of these intrinsic motivation sub scores was also analyzed. The Quiz score was calculated by how many correct answers they received on the quiz for each tutor.

This research seeks to explore whether sentiments are significant linear predictors of intrinsic motivation, flow, or information recall. As a result, my analysis used correlation, simple linear regression, and multiple linear regression analyses with a 5% level of significance to test my hypotheses. G*Power's Correlation: Bivariate Normal Model test was used to calculate correlation power values (Faul, Erdfelder, Lang, & Buchner, 2007).

Chapter 4: Results

The results below are organized by impact. I discuss sentiment impacts in the following order: pre-task topic net sentiment impact, pre-task CBI net sentiment impact, and post-task net sentiment impact. I then discuss other relevant findings such as positive sentiment vs negative sentiment impact, multiple linear regression results of post-task sentiments, topic sentiments vs factors of internalization, experience vs sentiment shift, and time on task relationships.

Table 4.1 summarizes the correlation and power value results for each type of sentiment (row labels) and the three learning outcomes (column labels). As discussed prior, note that net sentiment is positive sentiment minus negative sentiment. Correlation results summarized in Table 4.1 and regression results discussed in the following sections show strong support for hypotheses one and two, moderate support for hypothesis three and four, and no support for hypothesis five due to inconclusive data.

	Internalization	Flow Potential	Information Recall
Pre-task Topic Net Sentiment	0.65, >0.99**	0.56, >0.99**	0.06, 0.08
Pre-task CBI Net Sentiment	0.30, 0.72**	0.36, 0.88**	-0.13, 0.19
Post-task Topic Net Sentiment	0.71, >0.99**	0.60, >0.99**	0.01, 0.05
Post-task CBI Net Sentiment	0.68, >0.99**	0.60, >0.99**	0.02, 0.05
Pre-task Topic Positive Sentiment	0.55, >0.99**	0.48, >0.99**	-0.23, 0.48*
Pre-task CBI Positive Sentiment	0.31, 0.75**	0.29, 0.70**	-0.26, 0.58*
Post-task Topic Positive Sentiment	0.63, >0.99**	0.53, >0.99**	-0.02, 0.17
Post-task CBI Positive Sentiment	0.65, >0.99**	0.57, >0.99**	-0.06, 0.08
Pre-task Topic Negative Sentiment	0.14, 0.21	0.11, 0.15	-0.30, 0.72**
Pre-task CBI Negative Sentiment	0.07, 0.10	-0.04, 0.07	-0.20, 0.52*
Post-task Topic Negative Sentiment	-0.07, 0.10	-0.06, 0.09	-0.29, 0.83**
Post-task CBI Negative Sentiment	-0.09, 0.16	-0.11, 0.19	-0.26, 0.74**

Table 4.1: Summary of Correlation and Power Value Results. Pearson correlation coefficient is listed first. The power value of the correlation is listed after the comma and was calculated using a post-hoc analysis in G*Power. Note high (>0.70) power values (**) are green, moderate power values (*) are yellow, and low (<0.40) power values are red. Power values were calculated with the sample size of 70 and a 5% level of significance.

Pre-task Topic Net Sentiment Impact

This section shows the impact of pre-task topic net sentiment scores on internalization, flow potential, and information recall. Note that pre-task topic net sentiments are calculated by subtracting negative topic sentiment from positive topic sentiment. The pre-task label illustrates that the data was gathered from questionnaires taken before the tutors.

For each outcome, I illustrate findings in the context of the Camera Shots and Shogi tutors individually. After, I then discussed the results of the Camera Shots and Shogi tutor data combined into one simple linear regression (SLR) model. This order is used because there were significant differences in Camera Shot tutor and Shogi tutor results. For each SLR, the assumptions of equal variance, normality, independence of errors, and linearity were confirmed.

Effect on Internalization

For the Shogi tutor, a SLR test showed pre-task topic net sentiment was a predictor of internalization scores, $\beta = .22, t(68) = 4.21, p < .001$. Pre-task topic net sentiment also explained a moderate proportion of variance in internalization scores, $R^2 = .21, p < .001$. The two variables were moderately correlated, $r(68) = .46, p < .001$. The SLR model for the Camera Shots accounted for more variability in internalization scores than the SLR model for the Shogi tutor. The Camera Shots tutor SLR model showed pre-task topic net sentiment was a predictor of internalization scores, $\beta = .29, t(68) = 8.24, p < .001$. Pre-task topic net sentiment also explained a moderate proportion of variance in internalization scores, $R^2 = .50, p < .001$. The two variables had a moderate to strong correlation, $r(68) = .71, p < .001$.

Upon combining the data of the Camera Shots and Shogi tutors, I determined the total pre-task topic net sentiment and internalization scores. The total SLR model showed pre-task topic net sentiment was a predictor of internalization scores, *β = .28, t(68) = 7.05, p < .001*. Pre-task topic net sentiment also explained a moderate proportion of variance in internalization scores, *R^2 = .42, p < .001*. The two variables were moderately to highly correlated, *r(68) = .65, p < .001*. These results support the first hypothesis that pre-task topic net sentiment is a linear predictor of internalization.

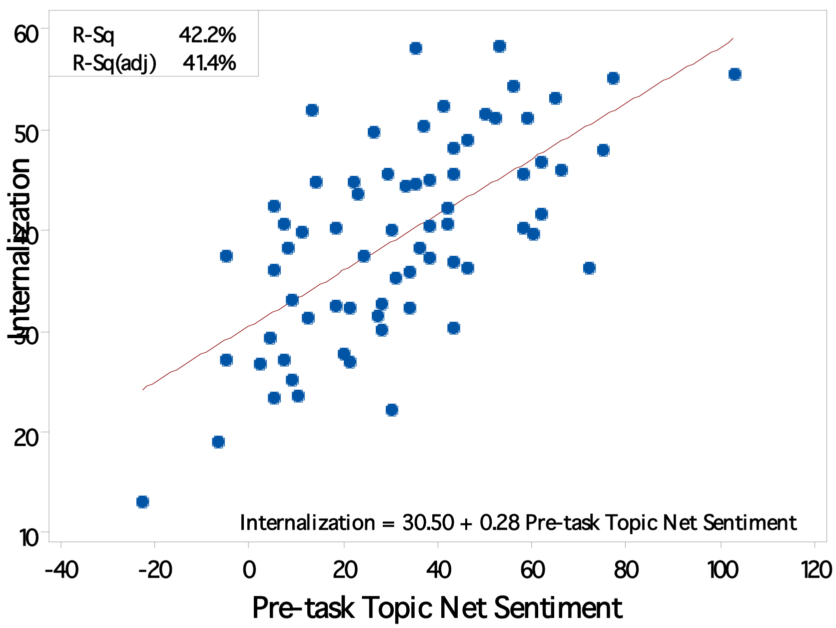

Figure 4.1: Total Pre-task Topic Net Sentiment vs Internalization

Effect on Flow Potential

The Shogi tutor SLR model showed pre-task topic net sentiment was a predictor of flow potential scores, *β = .44, t(68) = 4.69, p < .001*. Pre-task topic net sentiment explained a small proportion of variance in flow potential scores, *R^2 = .24, p < .001*. The two variables were moderately correlated, *r(68) = .49, p < .001*. For Camera Shots, the SLR model accounted for a similar amount of variability in flow potential. The Camera Shots tutor's SLR model showed pre-task topic net sentiment was a predictor of flow potential scores, *β = .36, t(68) = 4.74, p < .001*.

Pre-task topic net sentiment explained a small proportion of variance in flow potential scores, $R^2 = .25, p < .001$. The two variables were moderately correlated, $r(68) = .50, p < .001$.

With the data of the Camera Shots and Shogi tutors combined, I determined the total pre-task topic net sentiment and flow potential scores. The total SLR model showed pre-task topic net sentiment was a predictor of flow potential scores, $\beta = .47, t(68) = 5.63, p < .001$. Pre-task topic net sentiment also explained a low to moderate proportion of variance in flow potential scores, $R^2 = .32, p < .001$. The two variables were moderately correlated, $r(68) = .57, p < .001$. These results support the second hypothesis that pre-task topic net sentiment is a linear predictor of flow potential.

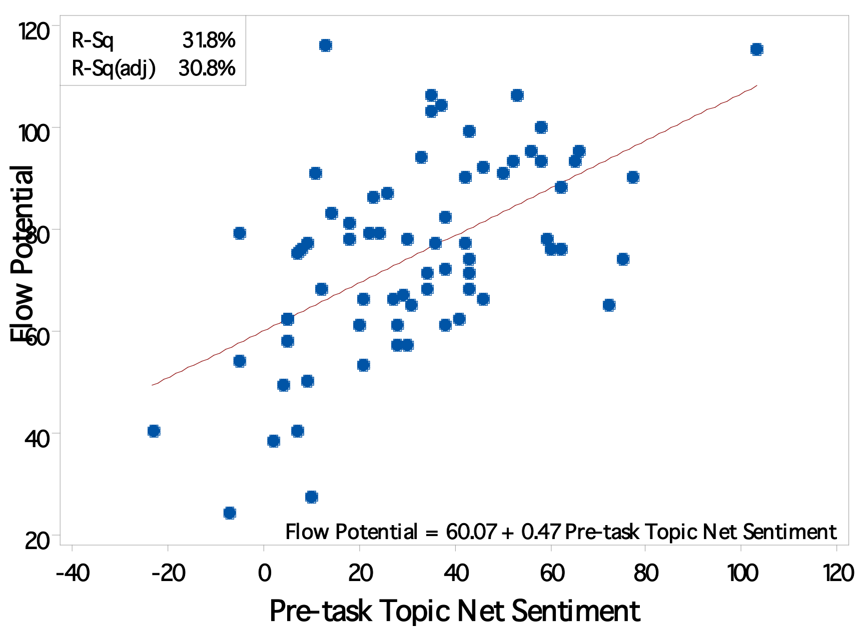

Figure 4.2: Total Pre-task Topic Net Sentiment vs Flow Potential

Effect on Information Recall

For the Shogi tutor, the SLR model showed pre-task topic net sentiment was not a predictor of information recall scores, $\beta = -.01$, $t(68) = -.44$, $p = .66$. The Camera Shots tutor's SLR model showed pre-task topic net sentiment was not a predictor of information recall scores as well, $\beta = .01$, $t(68) = .89$, $p = .38$. These results suggest that pre-task topic net sentiment is not a linear predictor of information recall.

Impact of Pre-task CBI Net Sentiment

Next, I will discuss the results of pre-task CBI net sentiment impact on internalization, flow potential, and information recall. Note that the pre-task label is because the data was collected before the volunteers took the tutors. As discussed prior, the CBI sentiment is the sentiment for method of learning. Like the prior section, I illustrate findings in the context of the Camera Shots and Shogi tutors individually. After, I then discuss the results of the Camera Shots and Shogi tutor data combined into one simple linear regression (SLR) model. This order is used again because there were significant differences in Camera Shot tutor and Shogi tutor results.

Effect on Internalization

For the Shogi tutor, a SLR test showed pre-task CBI net sentiment was not a predictor of internalization scores, $\beta = .09$, $t(68) = 1.37$, $p = .17$. Pre-task CBI net sentiment also explained an insignificant proportion of variance in internalization scores, $R^2 = .03$, $p = .17$. The two variables did not have a significant correlation, $r(68) = .16$, $p = .17$. The SLR model for the Camera Shots accounted for more variability in internalization scores than the SLR model for the

Shogi tutor. The Camera Shots tutor SLR model showed pre-task CBI net sentiment was a predictor of internalization scores, $\beta = .17, t(68) = 3.07, p = .003$. Pre-task CBI net sentiment also explained a small proportion of variance in internalization scores, $R^2 = .12, p = .003$. The two variables had a weak to moderate correlation, $r(68) = .35, p = .003$.

Upon combining the data of the Camera Shots and Shogi tutors, I determined the total pre-task CBI net sentiment and internalization scores. The total SLR model showed pre-task CBI net sentiment was a predictor of internalization scores, $\beta = .25, t(68) = 2.55, p = .01$. Pre-task CBI net sentiment also explained a small proportion of variance in internalization scores, $R^2 = .09, p = .01$. The two variables had a small to moderate correlation, $r(68) = .30, p = .01$. These results support the third hypothesis that pre-task CBI net sentiment is a linear predictor of internalization.

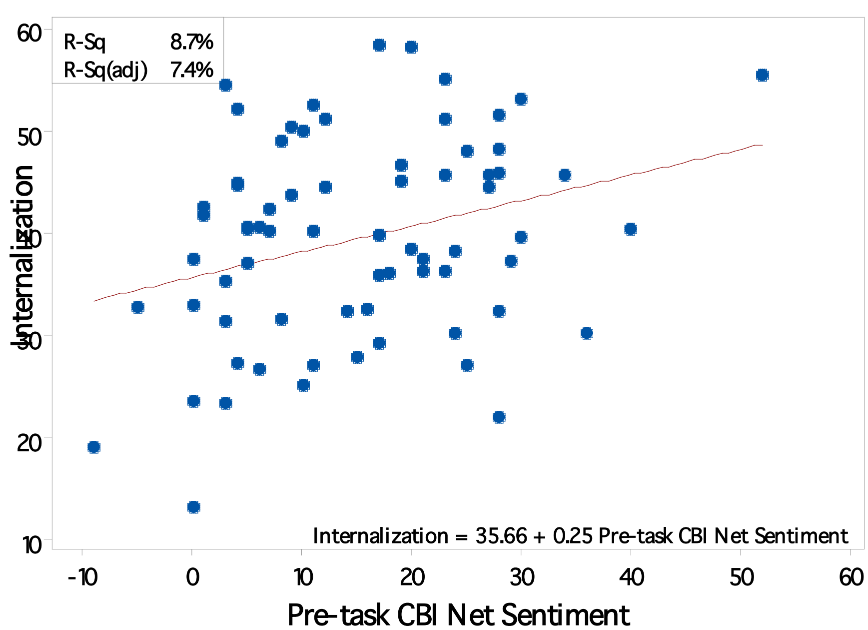

Figure 4.3: Total Pre-task CBI Net Sentiment vs Internalization

Effect on Flow

Pre-task CBI net sentiment explained more variance in flow potential scores than internalization scores. The Shogi tutor SLR model showed pre-task topic net sentiment was a predictor of flow potential scores, $\beta = .31, t(68) = 2.72, p = .008$. Pre-task CBI net sentiment explained a small proportion of variance in flow potential scores, $R^2 = .10, p = .008$. The two variables had a low to moderate correlation, $r(68) = .31, p = .008$. The Camera Shots tutor's SLR model accounted for a similar amount of variability in flow potential compared to the Shogi tutor. The Camera Shots tutor's SLR model showed pre-task CBI net sentiment was a predictor of flow potential scores, $\beta = .29, t(68) = 3.06, p = .003$. Pre-task CBI net sentiment explained a small proportion of variance in flow potential scores, $R^2 = .12, p = .003$. The two variables had a low to moderate correlation, $r(68) = .34, p < .001$.

With the data of the Camera Shots and Shogi tutors combined, I determined the total pre-task CBI net sentiment and flow potential scores. The total SLR model showed pre-task CBI net sentiment was a predictor of flow potential scores, $\beta = .60, t(68) = 3.22, p = .002$. Pre-task CBI net sentiment explained a low proportion of variance in flow potential scores, $R^2 = .13, p = .002$. The two variables had a low to moderate correlation, $r(68) = .36, p = .002$. These results support the fourth hypothesis that pre-task CBI net sentiment is a linear predictor of flow potential.

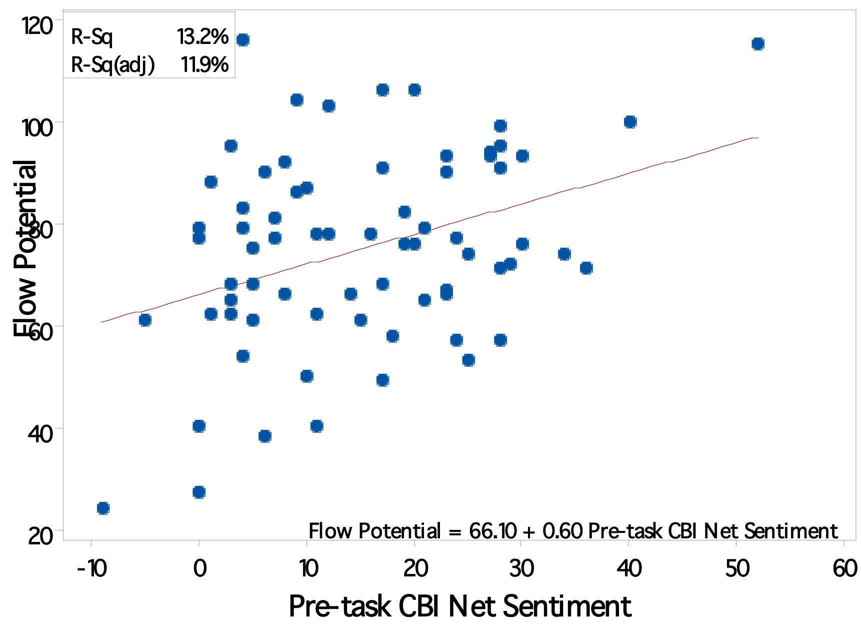

Figure 4.4: Total Pre-task CBI Net Sentiment vs Flow Potential

Effect on Information Recall

For the Shogi tutor, the SLR model showed pre-task CBI net sentiment was not a significant predictor of information recall scores at the 5% level of significance, β = *-.04, t(68) = -1.92, p = .06*. The Camera Shots tutor's SLR model provided further evidence that pre-task CBI net sentiment was not a predictor of information recall scores, β *< .01, t(68) = .16, p = .87*. These results suggest that pre-task topic net sentiment is not a linear predictor of information recall.

Impact of Post-Task Sentiments

Next, I will discuss the results for post-task sentiment impact on internalization, flow potential, and information recall. Note that the post-task label is because this data was collected after the volunteers had taken all the tutors. For each relationship discussed in this section, the Camera Shots and Shogi tutors showed similar results. Thus, the analysis in this section will focus only on total scores (combined Camera Shots tutor and Shogi tutor data). I will discuss the impacts of both post-task topic and post-task CBI sentiment in this section.

Effect on Internalization

Looking first at post-task topic net sentiment (Figure 4.5), the total SLR model showed post-task topic net sentiment was a predictor of internalization scores, $\beta = .28$, $t(68) = 8.41$, $p < .001$. Post-task topic net sentiment also explained a moderate proportion of variance in internalization scores, $R^2 = .51$, $p < .001$. The two variables were moderately to highly correlated, $r(68) = .71$, $p < .001$. Transitioning to post-task CBI sentiment scores, the total SLR model showed post-task CBI net sentiment was a predictor of internalization scores, $\beta = .48$, $t(68) = 7.67$, $p < .001$. Post-task CBI net sentiment also explained a moderate proportion of variance in internalization scores, $R^2 = .46$, $p < .001$. The two variables had a moderate to strong correlation, $r(68) = .68$, $p < .001$.

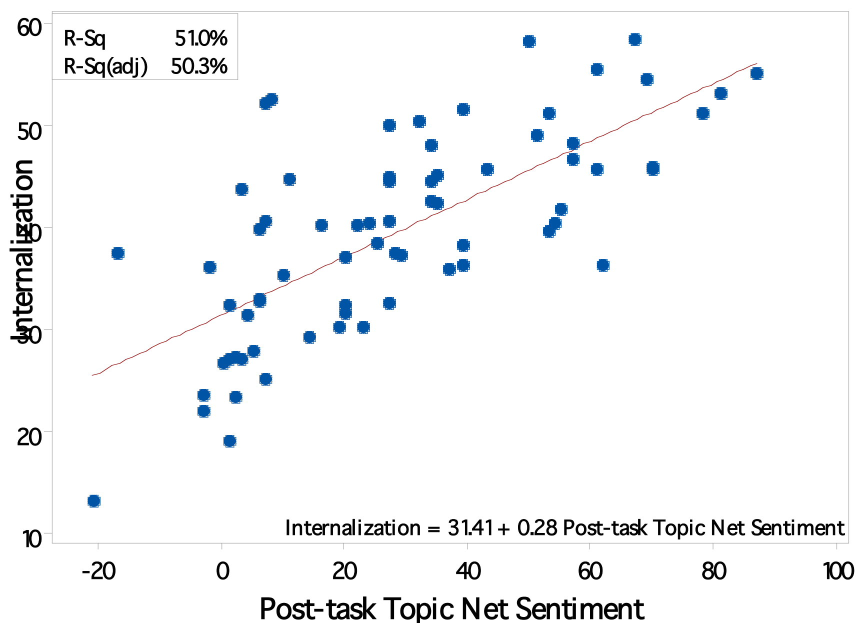

Figure 4.5: Total Post-task Topic Net Sentiment vs Internalization

Effect on Flow Potential

For post-task topic net sentiment, the total SLR model showed post-task topic net sentiment was a predictor of flow potential scores, $\beta = .46, t(68) = 6.12, p < .001$. Post-task topic net sentiment also explained a low to moderate proportion of variance in flow potential scores, $R^2 = .35, p < .001$. The two variables were moderately correlated, $r(68) = .59, p < .001$. For post-task CBI sentiment scores (Figure 4.6), the total SLR model showed post-task CBI net sentiment was a predictor of flow potential scores, $\beta = .83, t(68) = 6.25, p < .001$. Post-task CBI net sentiment also explained a low to moderate proportion of variance in flow potential scores, $R^2 = .36, p < .001$. The two variables had a moderate correlation, $r(68) = .60, p < .001$.

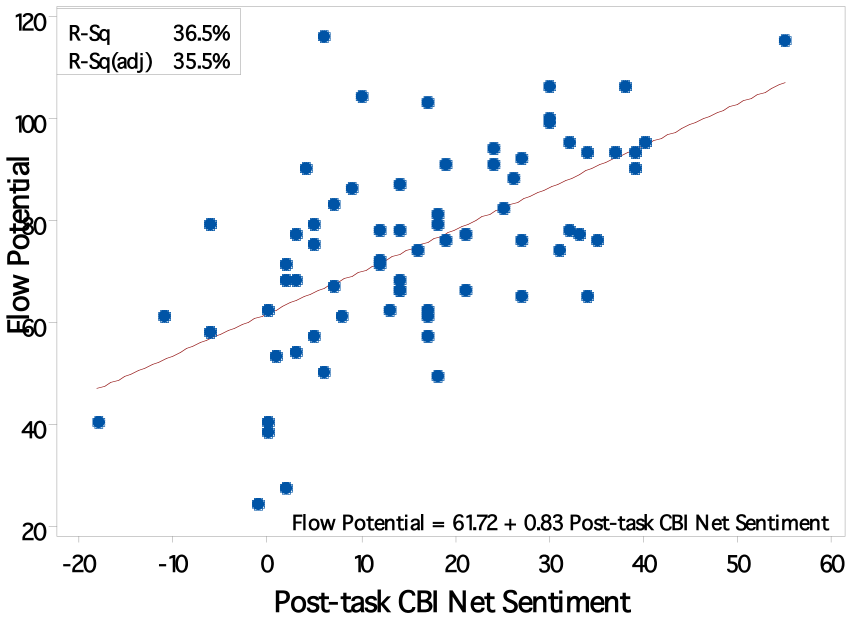

Figure 4.6: Total Post-task CBI Net Sentiment vs Flow Potential

Effect on Information Recall

Looking first at post-task topic net sentiment, the total SLR model showed post-task topic net sentiment was not a predictor of information recall scores, $\beta < .01$, *t(68) = .12, p = .90.* For post-task CBI net sentiment, the total SLR model showed post-task topic net sentiment was also not a predictor of information recall scores, $\beta = .01$, *t(68) = .57, p = .57.* Thus, neither post-task topic net sentiment nor post-task CBI net sentiment were linear of information recall.

Other Relevant Findings

Positive and Negative Sentiment Impact on Information Recall

As discussed above, results suggest that neither CBI net sentiment nor topic net sentiment had a linear relationship with information recall. As mentioned prior, information recall scores were calculated via two 10 question quizzes that contained multiple choice and true or false questions. The quizzes were similar in difficulty, but volunteers scored slightly worse overall on the Shogi quiz (Table 4.2).

Variable	***N***	***Mean***	***SE Mean***	***StDev***	***Min***	***Q1***	***Median***	***Q3***	***Max***
Camera Shots	70	6.14	0.22	1.81	2	5	6	7	10
Shogi	70	4.89	0.23	1.93	1	3.75	4	6.25	9

Table 4.2: Descriptive Statistics of Quiz Scores

To evaluate further, I looked at positive sentiment and negative sentiment independently to determine if either had a significant relationship with information recall scores. Highlighted in Table 4.1, the results suggest that negative sentiments had a significant linear relationship with information recall, and positive sentiments had no linear relationship with information recall. For instance, looking at a SLR test (Figure 4.7), the total SLR model showed pre-task topic negative sentiment was a predictor of information recall scores, $\beta = -.07, t(68) = -2.58, p = .01$. Pre-task topic negative sentiment explained a low proportion of variance in information recall scores, $R^2 = .09, p = .01$. The two variables had a low to moderate correlation, $r(68) = .30, p = .01$. Note that in addition to pre-task negative sentiment, post-task topic negative sentiment and post-task CBI negative sentiment had significant linear relationships with information recall as well (Table 4.1).

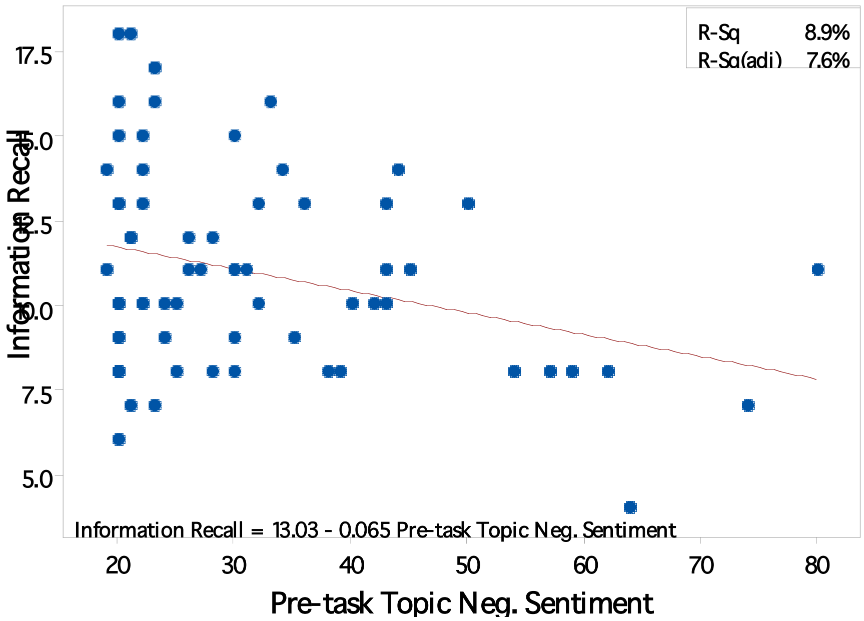

Figure 4.7: Pre-task Negative Sentiment vs Information Recall

Interestingly, although negative sentiment had a more significant relationship to information recall, it had a much weaker relationship with both internalization and flow potential scores (Table 4.1). Positive sentiment impacts were the opposite. Positive sentiments had a stronger linear relationship with internalization and flow potential and no significant linear relationship with information recall (Table 4.1).

Multiple Linear Regression (MLR) Analyses with Post-task Sentiments

Despite both post-task topic sentiment and post-task CBI sentiment being moderate positive linear predictors of internalization via individual SLR analyses, a MLR with both predictors barely increases the R^2 value. The MLR shows post-task CBI sentiment, with $p = .11$, as being a non-linear predictor at a 5% significance level in a model that includes post-task topic sentiment (Figure 4.8). The 3.69 variance inflation factors output suggests there is not an issue with multicollinearity.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	3637.7	1818.83	37.59	0.000
Post-task CBI Net Sentiment	1	128.6	128.62	2.66	0.108
Post-task Topic Net Sentiment	1	446.7	446.73	9.23	0.003
Error	67	3242.0	48.39		
Lack-of-Fit	63	2722.2	43.21	0.33	0.975
Pure Error	4	519.8	129.96		
Total	69	6879.7			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
6.95618	52.88%	51.47%	49.14%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	30.93	1.29	23.97	0.000	
Post-task CBI Net Sentiment	0.186	0.114	1.63	0.108	3.69
Post-task Topic Net Sentiment	0.1947	0.0641	3.04	0.003	3.69

Regression Equation

Internalization = 30.93 + 0.186 Post-task CBI Net Sentiment
+ 0.1947 Post-task Topic Net Sentiment

Figure 4.8: MLR Regression Test to Predict Internalization with Post-Task Sentiments

A MLR model predicting flow potential also fails to account for more variation in flow potential scores with both post-task net sentiments in the model. Post-task CBI net sentiment has a $p = .06$ while post-task topic net sentiment has a $p = .11$ (Figure 4.9). This is evidence that post-task CBI net sentiment may be more predictive of flow potential while post-task topic net sentiment may be more predictive of internalization. Again, the 3.69 variance inflation factors output suggests there is not an issue with multicollinearity.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	10006.7	5003.4	21.27	0.000
Post-task CBI Net Sentiment	1	863.2	863.2	3.67	0.060
Post-task Topic Net Sentiment	1	608.2	608.2	2.59	0.113
Error	67	15757.9	235.2		
Lack-of-Fit	63	13478.9	214.0	0.38	0.959
Pure Error	4	2279.0	569.8		
Total	69	25764.6			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
15.3360	38.84%	37.01%	33.67%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	60.88	2.85	21.40	0.000	
Post-task CBI Net Sentiment	0.481	0.251	1.92	0.060	3.69
Post-task Topic Net Sentiment	0.227	0.141	1.61	0.113	3.69

Regression Equation

Flow Score = 60.88 + 0.481 Post-task CBI Net Sentiment + 0.227 Post-task Topic Net Sentiment

Figure 4.9: MLR Regression Test to Predict Flow Potential with Post-Task Sentiments

A MLR model predicting information recall also fails to account for significantly more variation in information recall scores with both post-task negative sentiments in the model. Post-task CBI negative sentiment has a *p = .74* while post-task topic negative sentiment has a *p = .28* (Figure 4.10). This is evidence that post-task topic negative sentiment may be more predictive of information recall. The 3.21 variance inflation factors output suggests there is not an issue with multicollinearity.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	52.699	26.3497	3.03	0.055
Post-task CBI Neg. Sentiment	1	0.936	0.9357	0.11	0.744
Post-task Topic Neg. Sentiment	1	10.305	10.3051	1.18	0.280
Error	67	583.244	8.7051		
Lack-of-Fit	43	360.799	8.3907	0.91	0.622
Pure Error	24	222.444	9.2685		
Total	69	635.943			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.95045	8.29%	5.55%	0.55%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	13.023	0.887	14.69	0.000	
Post-task CBI Neg. Sentiment	-0.0302	0.0922	-0.33	0.744	3.21
Post-task Topic Neg. Sentiment	-0.0540	0.0497	-1.09	0.280	3.21

Regression Equation

Information Recall = 13.023 - 0.0302 Post-task CBI Neg. Sentiment
- 0.0540 Post-task Topic Neg. Sentiment

Figure 4.10: MLR Regression Test to Predict Information Recall with Post-Task Sentiments

Topic Sentiment Relationship with Internalization Components

Pre-task topic net sentiment had moderate correlation with each component of the internalization score. Interest had the strongest correlation with pre-task topic net sentiment $r(68) = .59, p < .001$, while perceived self-determination had the weakest correlation, $r(68) = .38, p < .001$. Note $p < .001$ for each correlation.

Post-task topic net sentiment saw similar results. Overall, post-task topic net sentiment had slightly higher correlations with each component of the internalization score. Again, interest had the strongest correlation with pre-task topic net sentiment, $r(68) = .64, p < .001$, while perceived self-determination had the weakest correlation, $r(68) = .59, p < .001$. Note $p < .001$ for each correlation.

	Pre-task Topic Sentiment	*Post-task Topic Sentiment*
Effort	.51	.58
Enjoyment	.57	.63
Interest	.59	.64
Perceived Importance	.58	.61
Perceived Self-determination	.38	.43

Table 4.3: Correlation Summary of Topic Sentiments vs Internalization Components. The Pearson correlation coefficient is listed for each relationship.

Change in Sentiment vs Experience

The results did not have enough experience score variety to accurately test this relationship. Although I expected Shogi and Camera Shots to be fairly low experienced topics, I expected CBI to have a wide range of experience levels to test the fifth hypothesis. This was not the case. Volunteers only had no experience to moderate experience.

Time on Task Relationships

I investigated the linear relationship between time on task and four outcomes. Those outcomes were post-task topic sentiment, internalization, flow potential, and information recall. Table 4.4 presents the findings. Note that time on task was measured in seconds and is how long the volunteer spent learning via the tutors. It does not include the time to take the questionnaires. For Shogi, the mean time on task was 240 seconds with a standard error of 22 seconds. For Camera Shots, the mean time on task was 253 seconds with a standard error of 32 seconds.

The results suggest that time on task and flow potential have a significant positive linear relationship for both tutors. Time on task and internalization have a significant positive linear relationship in the Shogi tutor but not the Camera Shots tutor at a 5% level of significance. Time on task and information recall and time on task and post-task topic net sentiment did not have a linear relationship.

	Shogi Seconds	***Camera Shots Seconds***
Post-task Topic Net Sentiment	.19	.19
	.13	.13
Internalization	.26	.23
	.03	.06
Flow Potential	.28	.25
	.03	.04
Information Recall	.13	.04
	.30	.72

Table 4.4: Time on Task Correlation Summary. The Pearson Correlation is listed first with the p-value below it.

Chapter 5: Discussion and Conclusion

Results showed that both pre-task and post-task net sentiments were positive linear predictors of internalization and flow potential. Neither topic nor CBI net sentiments were linear predictors of information recall. CBI net sentiment (method of learning sentiment) was a more effective predictor of flow potential, while topic net sentiment was a more effective predictor of internalization. Combining both CBI and topic net sentiments into a MLR model provided little improvement to the R^2 value compared to the SLR model. Results also showed a difference of impact between positive and negative sentiments. Negative sentiments had a more significant linear relationship with information recall while positive sentiments had a more significant linear relationship with internalization and flow potential.

Potential Explanations for Variability

This section focuses on providing potential causes of variability in the regression models. Five potential explanations are provided: Wrong expectations, self-regulation, other sentiments, extrinsic incentive, and distraction.

Wrong Expectation

I predict that the wrong learning expectation was the primary reason for the significant differences in pre-task topic net sentiment vs internalization findings between the Camera Shots and Shogi tutors. Most volunteers had little to no experience with Shogi, and I expect this led to misleading expectations of the Shogi learning experience and, thus, a large difference in pre-task and post-task sentiment scores. I assume the camera shots learning experience was easier to predict. This is why the pre-task and post-task sentiment scores were fairly similar.

Self-regulation

Regardless of sentiment scores, I predict a person with strong self-regulation ability would be more likely to be engaged and motivated than someone who is low in self-regulation. This research did not account for varying levels of self-regulation in volunteers, and I expect this was the primary cause of the variance in general. High self-regulators, for instance, can rationalize more personal value in tasks. For example, a high self-regulator could have rationalized learning about Shogi as an achievement-oriented task by convincing themselves that Shogi would be a beneficial interest on a resume.

Other Sentiments

This research accounted for sentiment for the learning topic and the learning medium (CBI). Other sentiments were not considered and likely contributed toward variance. For instance, I predict that the sentiment for the user interface, environment, and time of day could have contributed to engagement and motivation.

Extrinsic Incentive

Volunteers had the extrinsic incentive of extra credit to complete the study. This was needed to garner enough participants, but I expect it did affect the results. Research has shown that providing an extrinsic incentive for a task reduces an individual's intrinsic motivation to complete the task (Isen and Reeve, 2005). I predict intrinsic motivation scores would have been higher in a situation with no extrinsic motivator.

Distraction

Volunteers could complete the study at any location at any computer. Although volunteers were asked to complete the study in a non-distracting environment, this was not enforced. Environmental or online distractions likely affected intrinsic motivation and engagement.

Implications for ITSs

Based on the literature review and experiment results, I have developed four recommendations for ITS designers. Although this research was done in an ITS context, these recommendations apply to the more general CBI context as well.

Recommendation 1: Use emotion data to better understand learner intrinsic motivation, flow potential, and information recall.

Currently, emotion sensing technologies are being used primarily to predict affective states. Although understanding fleeting affective states offers advantages, long-term emotional associations (sentiments) should not be ignored. Post-task sentiment results suggest that sentiment measurements could account for about 51% of internalization variation, about 36% of flow potential variation, and about 9% of information recall variation. I predict post-task sentiment estimations can be calculated through a combination of self-report measures and emotion sensing technology measures. I recommend taking advantage of this opportunity to better understand the learner's motivation, engagement, and information recall patterns by using emotion data to estimate learner sentiments.

Recommendation 2: Select a Sentiment Measurement Strategy that Fits the Tutor's Purpose

Results showed that different sentiments had stronger relationships with different learning outcomes. Results suggest that topic net sentiments had the strongest relationship with internalization, CBI net sentiments had the strongest relationship with flow potential, and topic negative sentiments had the strongest relationship with information recall. Thus, ITS designers should align ITS measurement strategies to match the tutor's objective. For example, a one-lesson ITS that prioritizes information recall should focus on measuring negative sentiments. However, a multi-lesson ITS that requires user motivation to progress through all the lessons should prioritize topic net sentiment measurements.

Recommendation 3: Account for the impact of prior sentiments

Pre-task sentiment results suggest that about 40% of internalization variation, 30% of flow potential variation, and about 9% of information recall variation could be explained by preexisting topic sentiments held by the learner. Due to the impactful nature of these predetermined sentiments, it is important for ITSs to understand and account for these predetermined sentiments. ITSs should seek to determine what preexisting sentiments the learner has prior to engaging with the lesson and monitor how those sentiments change through the learner's ITS experience.

Recommendation 4: Sentiment shift should be a key metric to monitor to avoid a negative sentiment spiral or to create a positive sentiment spiral

As discussed prior, I assume that sentiments contribute to attitude formation because of their similarity to the affective component of attitude. If this is true, then a decline in task sentiment will increase the chance of negative emotion appraisals for future events surrounding that task (Figure 2.3). In other words, having a negatively valenced experience increases the chance to have another negative experience. I call this the negative sentiment spiral. Research suggests this spiral effect can work in the positive direction as well. Positive sentiments will contribute to positive attitude formation, and the results also suggest that sentiments are predictive of intrinsic motivation and flow. As discussed prior, both intrinsically motivated states and flow states have been shown to increase positive emotions. In other words, if there is a shift toward a positive task sentiment, there will be a greater chance to appraise positive emotions for future events surrounding that task. I call this the positive sentiment spiral. Sentiment shift monitoring can help enable an ITS to prevent a negative sentiment spiral or trigger a positive sentiment spiral.

Recommendations for Future D2P Development

This research involved building two tutors using the D2P program. Throughout tutor development, I identified potential usability enhancements for the D2P program that could potentially help future D2P tutor designers. I recommend: Increasing question layout flexibility, adding duplication capability, enhancing the learner data export feature, and enabling tutor access without sign in.

Increase Question Layout Flexibility

The PANAS questionnaires included 20 questions structured on a single question set. D2P has a default layout for how questions within a question set are laid out on the page. This default layout has questions in two columns with the questions fairly close together. For this research, this was not an ideal structure as it brought complaints from pilot participants. It was easy for volunteers to accidentally skip over a question or be overwhelmed by the amount of text that was close together. As a result, the first recommendation is to increase question layout flexibility for question sets.

Add Duplication Capability

In this research study, I had duplicate questionnaires. For instance, a slightly different version of PANAS was taken 9 times by each participant. The lack of duplication capability for individual questions, question sets, and quizzes forced us to build similar question sets multiple times. A duplication feature for questions, question sets, and quizzes would save future tutor designers a significant amount of time if they have similar questionnaires.

Enhance Learner Data Export

Most of the questions did not have correct answers because this research concerned how the participant answered the question. Currently, the learner data export feature does not show how volunteers answered questions. It only shows the correctness and the time of the answer. As a result, I had to collaborate with a database administrator each time I wanted to analyze new participant data. An enhanced export feature would facilitate data analysis for any future questionnaires.

Enable Access Without Sign in Option

The D2P program requires every user to sign into a D2P account or to create an account. This requirement could be a challenge for wide-spread tutor use. Although volunteers could create their own account and sign, I felt this could be confusing. To give participants access to the tutors, I created a D2P account for every user and provided them an email and password. I recommend exploring a feature that allows anyone with a link to access a tutor. This change would assist future designers that want a broad audience of users.

Areas for Future Research

Although this research provided insight, it also created more questions. This section focuses on potential areas for future research.

Sentiment and Emotion Sensing Technologies

This research calculated sentiment using self-report measures. A study that used an emotion sensing tool such as Affectiva to calculate sentiments from emotional valences would be more accurate than the modified PANAS questionnaire. Additionally, findings would be more applicable to ITS designers using GIFT.

More Experienced Tasks

I used the learning topics of Camera Shots and Shogi to explore the relationship between topic sentiment and learning outcomes. Most volunteers had fairly low experience with both of these topics. I predict a stronger relationship between sentiments and learning outcomes will exist for tasks that learners have significant experience in. This needs to be validated, however.

Impact of Other Sentiments

In this research, I explored topic sentiment and learning medium sentiment. I predict there are other sentiments that are relevant to motivation and engagement. These include environment sentiment, user interface sentiment, social situation sentiment, and time sentiment. Additionally, I expect topic sentiment and learning medium sentiment can be broken down into more specific sentiments. All of these potential sentiments can and should be investigated.

Sentiment and Attitudes

Based on the nearly identical nature of sentiment and affective attitude, I assume that sentiment contributes to attitude formation. However, validation and exploration in this area is required. In addition, the investigation into how the cognitive component of attitude influences sentiment could provide useful insight for GIFT's ABC User Model.

Concluding Remarks

This research blended insights from ITSs, emotion and learning, and motivation and engagement to explore how sentiments relate to key learning outcomes. Results showed significant linear relationships between sentiments and internalization and between sentiments and flow potential. The findings also suggest that positive and negative sentiments have different impacts on key learning outcomes with negative sentiment having a significant linear relationship with information recall. I hope ITS designers use this research to improve motivation-adaptive learning in ITSs. Moreover, I hope these findings illustrate the importance of measuring and calculating long-term emotional associations to optimize the computer-based learning experience.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 8–918.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*(42), 167–207.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In *Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (pp. 17–24). Amsterdam.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, *13*(2), 111–127.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, *21*(4), 755–765.

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: from the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87–122.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 602–642). New York:

Cambridge University Press.

Brave, S., & Nass, C. (2009). Emotion in human-computer interaction. In A. Sears & J. A. Jacko (Eds.), *Human-Computer Interaction Fundamentals* (1st ed., pp. 54–61). Boca Raton: Taylor & Francis Group, LLC.

Breckler, S. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*(6), 1191–1205.

Challco, G. C., Andrade, F. R. H., Borges, S. S., Bittencourt, I. I., & Isotani, S. (2016). Toward a unified modeling of learner's growth process and flow theory. *Educational Technology & Society*, *19*(2), 215–227.

Choi, D. H., Kim, J., & Kim, S. H. (2007). ERP training with a web-based electronic learning system: The flow theory perspective. *International J.ournal of Human-Computer Studies*, *65*, 223–243.

Cohen, P. A., & Dacanay, L. S. (1992). Computer-based instruction and health professions education: A meta-analysis of outcomes. *Evaluation & The Health Professions*, *15*(3), 259–281.

Connor, M., Sheeran, P., Godin, G., & Germain, M. (2013). Some feelings are more important: Cognitive attitudes, affective Attitudes, anticipated affect, and blood donation. *Health Psychology*, *32*(3), 264–272.

Corbalan, G., Kester, L., Jeroen, &, Van, J. G., & Nboer, M. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science*, *34*(1), 399–422.

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct. *The British Journal of Clinical Psychology*, *43*(1), 245–265.

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

Csikszentmihalyi, M., & Lefevre, J. (1989). Optimal experience in work and leisure. *Journal of*

Personality and Social Psychology, *56*(5), 815–822.

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personal Ity and Social Psychology*, *59*(2), 202–216.

Erez, A., & Isen, A. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology*, *87*(6), 1055–1067.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fletcher, J. D. (2017). The digital tutoring opportunity. In *Building America's Skilled Technical Workforce: The Role of Digital Tutors*. Washington, DC.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research Spring*, *74*(1), 59–109.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA Handbook of Educational Psychology* (3rd ed., pp. 451–473). Washington, DC: American Psychological Association.

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review*, *3*(1), 5–39.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*(4), 549–571.

Hughes, M. G., Day, E. A., Wang, X., Schuelke, M. J., Arsenault, M. L., Harkrider, L. N., & Cooper, O. D. (2013). Learner-controlled practice difficulty in the training of a complex task: cognitive and motivational mechanisms. *Journal of Applied Psychology*, *98*(1), 80–98.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, *29*(4), 297–325.

Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport and Exercise Psychology*, *30*(1), 561–587.

Jang, H., Reeve, J., Ryan, R. M., & Kim, A. (2009). Can self-determination theory explain what underlies the productive, satisfying learning experiences of collectivistically oriented korean students? *Journal of Educational Psychology*, *101*(3), 644–661.

Kavanagh, D. J., & Bower, G. H. (1985). Mood and self-efficacy: Impact of joy sadness on perceived capabilities. *Cognitive Therapy and Research*, *9*(5), 507–525.

Keller, J., & Bless, H. (2008). Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. *Personality and Social Psychology Bulletin*, *34*(1), 196–209.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78.

Lazowski, R., & Hulleman, C. (2015). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918.

Mageau, G. A., & Vallerand, R. J. (2007). The moderating effect of passion on the relation between activity engagement and positive affect. *Motivation and Emotion*, *31*(1), 312–321.

Martin, F., Klein, J. D., & Sullivan, H. (2007). The impact of instructional elements in computer-based instruction. *British Journal of Educational Technology*, *38*(4), 623–636.

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

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology* (pp. 239–263). Springer Netherlands.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, *7*(2), 133–144.

Ortony, A., Clore, G., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Panksepp, J. (2003). At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition*, *52*(1), 4–14.

Picard, R. (2006). Building an affective learning companion. In *Intelligent Tutoring Systems* (pp. 811–811). Berlin.

Reeve, J., Jang, H., Hardre, P., & Omura, M. (2002). Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. *Motivation and Emotion*, *26*(3), 183–207.

Ritter, F. E., Yeh, K.-C., Cohen, M. A., Weyhrauch, P., Kim, J. W., & Nicholas Hobbs, J. (2013). Declarative to procedural Tutors: A family of cognitive architecture-based tutors. In *Proceedings of the 22nd Annual Conference on Behavior Representation in Modeling & Simulation*. Ottawa: BRIMS Society.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: theory, methods, research*. New York: Oxford University Press.

Ryan, R. M. (1995). Psychological needs and the facilitation of integrative processes. *Journal of Personality*, *63*(3), 397–427.

Ryan, R. M., Bernstein, J., & Warren Brown, K. (2010). Weekends, work, and well-being: psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. *Journal of Social and Clinical Psychology*, *29*(1), 95–122.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In R. M. Ryan & E. L. Deci (Eds.), *Handbook of Self-Determination Research* (pp. 3–33). Rochester: University of Rochester Press.

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

Shernoff, D. J., & Csikszentmihalyi, M. (2009). Flow in schools cultivating engaged learners and optimal learning environments. In *Handbook of Positive Psychology in Schools* (pp. 131–145). New York: Routledge.

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*. American Educational Research Association.

Sottilare, R., Brawner, K., Goldberg, B., & Holden, H. (2012). *The generalized intelligent framework for tutoring (GIFT)*. Orlando.

Sottilare, R., Graesser, A., Hu, X., & Holden, H. (2013). *Design recommendations for intelligent tutoring systems*. (R. Sottilare, A. Graesser, X. Hu, & H. Holden, Eds.). Orlando: U.S. Army Research Laboratory.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*(4), 970–987.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331–347.

Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, *100*(2), 460–472.

Turner, J. C., Meyer, D. K., Midgley, C., & Patrick, H. (2003). Teacher discourse and sixth graders' reported affect and achievement behaviors in two high-mastery/high-performance mathematics classrooms high-mastery/high-performance mathematics classrooms. *The Elementary School Journal*, *103*(4), 357–382.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

Wan, C.-S., & Chiou, W.-B. (2006). Psychological motives and online games addiction: A test of flow theory and humanistic needs theory for taiwanese adolescents. *Cyberpsychology & Behavior*, *9*(3).

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2008). Development of achievement motivation. In W. Damon, R. M. Lerner, D. Kuhn, R. S. Siegler, & N. Eisenberg (Eds.), *Child and Adolescent Development*. Hoboken: John Wiley & Sons, Inc.

Woolf, B. P., Arroyo, I., Cooper, D., Burleson, W., & Muldner, K. (2010). Affective tutors: Automatic detection of and response to student emotion. In R. Nkambou, R. Mizoguchi, & J. Bourdeau (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 207–227). Springer Science & Business Media.

Woolf, B. P., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognising and responding to student affect. *Learning Technology*, *4*(3), 128–164.