Tracking Engagement: A Machine Learning Framework for Estimating Affective Engagement

By

Aramael Andres Peña-Alcántara

B.S. Drama and Theatre Arts with a Concentration in Design,
Columbia College, Columbia University, 2017
B.S. Mechanical Engineering, The Fu Foundation School of
Engineering and Applied Science, Columbia University, 2017

SUBMITTED TO THE CIVIL AND ENVIRONMENTAL ENGINEERING DEPARTMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN CIVIL AND ENVIRONMENTAL ENGINEERING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author: ______________________________

Civil and Environmental Engineering
May 15, 2020

Certified by: ______________________________

Dr. John Williams, Thesis Supervisor
Professor, Civil and Environmental Engineering

Accepted by: ______________________________

Dr. Colette L. Heald,
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

This Page Left Intentionally Blank

Tracking Engagement: A Machine Learning Framework for Estimating Affective Engagement

By

Aramael Andres Peña-Alcántara

Submitted to the Civil and Environmental Engineering Department on May 15, 2020, in partial fulfillment of the requirements for the degree of Master of Science in Civil and Environmental Engineering

ABSTRACT

Globally, construction fatality counts remain among the highest of all industries. As part of efforts to improve workers occupational health and safety, most companies provide workers with ongoing safety training. Yet accidents continue to take place, as there is a lack of understanding on how to increase the knowledge transfer that would help improve safety. The goal of this thesis is to automate and improve manual observation methods, presently used to determine construction workers' engagement during training courses by applying machine learning techniques to video images. This thesis proposes a framework to measure construction workers' engagement during training courses by unobtrusively analyzing engagement through body and pose estimation, codifying who is speaking and understating the predicted emotional state of a given worker through their facial expressions of emotion at specific lectures times through state-of-the-art computer vision techniques. The framework was prototyped on fifteen graduate and undergraduate students from a private university in the United States during four class sessions in a stadium set up classroom by three high definition cameras. The proposed system can enhance our understanding of learning processes within classroom contexts, while reducing the labor-intensive process of traditional observations methods, and allowing for the observation of a full class simultaneously. Further, the repeatability and standardization of objective observations will be improved as it will no longer depend on the skills of the observer and on his or her ability to capture and make sense of what was observed.

Thesis Supervisor: John Williams
Title: Professor of Civil and Environmental Engineering

Acknowledgments

"Bernard of Chartres used to compare us to dwarfs perched on the shoulders of giants. He pointed out that we see more and farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature."

—**John of Salisbury**, Metalogicon (1159)

"I say with Didacus Stella, a dwarf standing on the shoulders of a giant may see farther than a giant himself."

— **Robert Burton**, The Anatomy of melancholy (1621)

"If I have seen further it is by standing on the shoulders of Giants."

— **Isaac Newton**, Letter to Robert Hooke (1675)

Sincere and grateful acknowledgements to those predecessors. Special thanks to all my friends and family — in the broad, Dominican sense of the words.

This thesis is dedicated to my parents, for they saw in me something I had yet to see in myself.

Table of Contents

Table of Figures

Table of Tables

1. Introduction

This thesis proposes a framework that would measure construction workers' engagement during construction training courses conducted by unobtrusively analyzing cues from the worker's face, body posture, and hand gestures through state-of-the-art computer vision techniques. An initial prototype of the implementation of this framework has been created and tested in a classroom environment at MIT. This tool would allow for an analysis of affective engagement to validate and design new safety trainings. In the US, construction fatality counts remain among the highest of all industries, accounting for nearly 19% of all workplace fatalities in 2017 (Bureau of Labor Statistics, 2017). Globally, construction's accidental death and injury rate is more than two times higher than the average of other industries (Sousa et al., 2015). To stem this issue and to improve worker's occupational safety and health (OSH), most companies provide workers with ongoing safety training. The basic idea of safety training is to impart safety knowledge (an individual's knowledge of how to perform work safely) and safety motivation (an individual's willingness to exert effort to enact safe behaviors) to ensure that workers know how to work safely and why safety is imperative to their wellbeing (Christian et al., 2009). Yet, while these training have reduced the number of casualties in the industry, still, accidents continue to take place, as there is not a clear understanding of what training features increase the knowledge transfer from trainer to trainee that would help improve safety (Hale, 1984). It is particularly important to understand which learning experiences lead to training transfer, defined as the ability to extend what has been learned in training to their workplace (Byrnes, 2007).

2. Motivation

Occupational injuries affect workers, employers, and society writ large through its impact on medical costs, workplace productivity, and the pain and suffering caused by injuries. Employers, in particular, experience work disruptions and costs associated with workplace injuries. To achieve productivity goals, employers frequently train workers in a variety of work practices to reduce injuries (Waehrer & Miller, 2009). Employers invest in designing, developing, and delivering training programs to equip workers with the skills necessary to recognize and manage hazards in complex environments (Hinze Jimmie & Gambatese John, 2003). A national survey found establishments with 50 or more employees paid $13.2 billion to training staff in 1994, $237 per employee respectively (Waehrer & Miller, 2009). Despite these training efforts, more than 70% of accidents in construction projects are associated with poor safety knowledge (Haslam et al., 2005). This reveals that the improvements to training programs are not necessary and sufficient to reach many firm's desirable levels of hazard recognition in practice, and the expected return on investment for these trainings has not been achieved (Namian Mostafa et al., 2016).

3. Literature Review

Within early industrial accident research literature, there are almost as many studies that show that job training had no effect on safety as studies which showed a positive effect. Yet when controlling for training programs which were conducted with "due regard to the principles of good training design" there exist significant reductions in accidents in trained groups as opposed to the untrained (Van Zelst, 1954). The shortcomings of early research not showing any effect on safety has been generally attributed to weaknesses in training programs, including improper delivery and inferior materials (Hale, 1984). It is imperative to address these shortcomings as the difference between effective and ineffective training is death, injury, pain, suffering, and lost profits (Robotham, 2001). Training effectiveness is best explored by posing three evaluative questions to understand training effectiveness (Burke et al., 2006; Cohen & Cooligan, 1998):

— **Does training increase safety knowledge?** Instructors must determine if participants successfully met the training objectives. This can be measured by pre- and post- test or performance demonstrations

— **Does training result in safer workplace behavior?** Managers should observe if there has been a successful transferring of learning from the classroom to the workplace. Often this is measured by job observation and job performance surveys.

— **Does training result in better safety outcomes?** If good training has been delivered effectively to the correct individuals, consistent application should result in better job performance and fewer safety incidents.

Workplace training is designed to educate adults from various backgrounds who will face different challenges (Wilkins, 2011). More recently, the amount of improvement and retained improvement was found tied to the training's level of engagement (Taylor, 2015). Less engaging formats included lectures, videos, and pamphlets, while hands-on exercises which allowed for development of knowledge in stages were considered the most engaging. The most engaging methods of training produced approximately three times the improvement as the least engaging, and the more engaging forms produced better safety outcomes (Burke et al., 2006).

While the term is used without a clear definition in the *Burke et al*. meta-review, it may be more useful to consider a theoretical model for the term. Student engagement is largely understood to be a meta-construct developed and primarily utilized by academic educational researchers. Conceptualized in the 1980s with an emphasis on reducing student alienation, boredom, and dropout (J. Finn & Zimmer, 2012). In both popular and research definitions, the term is seen to encapsulate the qualities that are necessary for today's students. The term engagement is multifaceted and is best understood as a multidimensional construct that unites multiple components in a meaningful way (Fredricks et al., 2004). Many models of engagement have been presented in the literature, often utilizing different terminology, but there are four dimensions that appear repeatedly: academic, social, cognitive, and affective engagement *Academic engagement* refers to exhibited behaviors related directly to the learning process (i.e. attentiveness, and completing assignments). There exists a minimum "threshold" levels of academic engagement that is considered essential for any learning to occur (Appleton et al., 2006; J. Finn & Zimmer, 2012; Fredricks et al., 2004; Jimerson et al., 2003; Rumberger & Lim, 2008).

- *Social engagement* measures the extent to which students follow written and unwritten classroom expectations (i.e. attendance, timeliness, positive instructor and peer interactions). This requires physical and mental presence; and not withdrawing from participation or actively disrupting others work.
- *Cognitive engagement* is the expenditure of effort needed to comprehend complex ideas (i.e. asking questions, clarifying concepts, reading additional material, and reviewing material beyond a minimal understanding). High levels of cognitive engagement facilitate student's learning of complex material, and employs the use of self-regulation and other cognitive strategies to guide learning.
- *Affective engagement* is the emotional response characterized by feelings of involvement (J. Finn & Zimmer, 2012). This form of engagement represents students' feelings of education as a set of activities worth pursuing and provides an incentive to continue to persist in educational endeavors.

The emotions included in the definitions of affective engagement duplicate an earlier body of work on achievement emotions (AEs), which have been found to affect the way students learn and increase their academic achievement when performing tests, assignments, and/or performances in both in-class and take-home tasks (Peterson et al., 2015). Research into AEs has been dominated by Control-Value Theory (CVT) which models the effects of emotions on learning and performance (Pekrun, 2006). In the CVT model, emotions are tied directly to achievement activities or achievement outcomes. Achievement is defined as the quality of activities or outcomes when evaluated by some standard of excellence (Heckhausen, 1991). Before CVT, research on AEs mostly focused on the emotions relating to outcomes (e.g., the joy and pride experienced by students when meeting their academic goals, or the frustration and shame they feel when their efforts fail). CVT expanded the theory by positing that emotions pertaining to accomplishing achievement activities are also AEs (e.g. the excitement of learning, boredom experienced with curriculum instruction, or anger when having to perform numerous assignments). In the end, the difference of activity vs. outcome emotions pertains to the focus of AEs (Pekrun et al., 2007). The theoretical work on AEs outlines finer distinctions than are present in engagement literature. CVT proposes that AEs have three dimensions: valence (positive vs. negative; or pleasant vs. unpleasant), the degree of activation, and the object focus. Using these three dimensions, AEs can be organized in a three-dimensional taxonomy as seen in **Table 1** above (Pekrun et al., 2007).

***Table 1** A Three-Dimensional Taxonomy of Achievement Emotions*

	Positive[a]		Negative[b]	
Object Focus	Activating	Deactivating	Activating	Deactivating
Activity Focus	Enjoyment	Relaxation	Anger Frustration	Boredom
Outcome Focus	Joy Hope Pride Gratitude	Contentment Relief	Anxiety Shame Anger	Sadness Disappointment Hopelessness

__a__ Positive, pleasant emotion; __b__ Negative, unpleasant emotion; Reproduced from (Pekrun et al., 2007)

Valence, within the CVT model, scales positive (i.e. joy, hope, pride, gratitude, contentment, relaxation, and relief) vs. negative emotions (i.e. anger, frustration, anxiety, shame, anger, boredom, sadness, disappointment, and hopelessness). The effect of AEs can be considered to be activating, when it contributes to greater learning effort, neutral towards learning, or deactivating when it is maladaptive towards learning. Taken together these three-dimensional taxonomies interact so that emotions might be positive and deactivating and negative and activating, as shown in **Table 1** above. Lastly AEs are context specific and exist in relation to a learning activity or to an outcome. This temporal understating of AEs allows the model to capture the variation in AEs leading up to, during, and after an assessment activity and in relation to its learning outcome (Pekrun, 2006). This model was furthered with the now popular Academic Emotions Questionnaire (AEQ) which focuses on 9 emotions (4 positive AEs: enjoyment, hope, pride, relief; and five negative AEs: anger, anxiety, shame, hopelessness, and boredom) (Pekrun et al., 2002).

As varied as the definitions of student engagement are, there is a multitude of methods for measuring engagement (Fredricks et al., 2004). The most common method for determining student engagement is the student self-report survey measures (Fredricks & McColskey, 2012). In these surveys, students are provided with statements reflecting various aspects of engagement and they are to select the response they most identify with. This method is widely used given its practicality, low cost and easiness to administer in classroom settings. However, several concerns with self-report measures includes lack of student honesty under certain conditions (e.g., no anonymity provided) (Appleton et al., 2006; Garcia & Pintrich, 1996); broadly measured items (e.g., I enjoyed the lectures) as opposed to items worded to reflect engagement in specific tasks and situations. For researchers interested in studying how affective engagement varies as a function of contextual factors, affective activities and outcomes, these general items may not be appropriate.

Another method for assessing student engagement is instructor report on students. Teachers' rating scales include items assessing both behavioral and emotional engagement (Skinner & Belmont, 1993), multidimensional models of engagement (i.e., behavioral, emotional, and cognitive) (Wigfield et al., 2008), teacher ratings of student participation as indicative of behavioral engagement (J. D. Finn et al., 1991, 1995) and teacher ratings of adjustment to school, as indicative of engagement (Birch & Ladd, 1997; Buhs & Ladd, 2001). Some of the shortcomings of these ratings is the ability of the teacher to properly interpret the affective engagement of the student as well as their ability to assess all students given classroom size. Other researchers have used structured and semi-structured interview techniques to assess engagement and provide insight into the reasons for variability in levels of engagement. While interviews are beneficial as they can provide detailed accounts of how students interpret their school experiences, and how these experiences relate to their engagement (Fredricks et al., 2004), they are not without problems. The skills and biases of the interviewer could potentially impact the quality, depth, and interviewee's response, as well the reliability (stability and consistency) and validity of interview findings (McCaslin & Good, 1996).

Lastly observational methods at both individual and classroom level have been used to measure engagement (Fredricks & McColskey, 2012). Observational methods primarily utilize individuals' students on and off task behavior—the period of time during which a student is actively engaged in a learning activity— as an indicator of academic engagement. A meta-analysis of the state of the research emphasizes that the effects of academic and social engagement on educational accomplishments are consistently statistically significant and moderate to strong, and the effect on affective engagement on academic engagement and completion rates are consistently positive (J. Finn & Zimmer, 2012). This understanding further emphasizes the value of observational measures as they best provide an objective ground truth which most corresponds with academic accomplishments (J. Finn & Zimmer, 2012; Peterson et al., 2015). Yet these observational measures are labor intensive and usually involve only a small number of students and contexts for observation, raising concerns about the generalizability to other settings. Lastly the quality of observations depends heavily on the skills of an observer on his or her ability to capture and make sense of what was observed (Turner & Meyer, 2000). Training observers is no small task, one student observational methodology, the Behavioral Observation of Students in Schools, states that about 10–15 hours of training is required to become proficient at administering the measure (Fredricks & McColskey, 2012).

The goal of this thesis is to reduce the inefficiencies (e.g. observation training lack of knowledge, skills, or biases in observers), cost of a manual observation methods (e.g. costs to train observers) to determine construction workers' engagement during construction training courses to further minimize construction accidents. The proposed new automated camera vision techniques would measure emotional states using machine learning algorithms. The proposed system posits that this camera based observational measure will perform as well as observational experts as it would both understand academic engagement through body and pose estimation, social engagement through codifying who is speaking, and affective engagement by understating the predicted emotional state of a given participant through their facial expressions of emotion (FEE).

4. Theory

4.1. Theoretical Focus of Attention Model

When a word is spoken, all who happen to be in perceptual range of this utterance will have some sort of participation status relative to it (Goffman, 1981). To display the relationship among these participants, a participation framework model emerges, changes, and adapts as the interaction between speakers and hearers. Participants take on their status in a speaker or hearer role and assumes their places in the participation framework for each utterance.

Considering the complex structure of a polyadic interaction in general, one has to first consider that a person who delivers an utterance does not necessarily address it to everybody in the acoustic reach of his or her voice. Thus, there are participants who are directly "addressed" by a speaker and others who are not involved but still within earshot. This creates a distinction between ratified (direct participation) and unratified (not direct participation) participants. The hearer to whom the speaker allocates the right to take over at the next turn is given the title of "conversational partner." This partner is associated with the obligation of a high level of attentiveness. Those who are also addressed but not assumed to be the next speaker but may later be involved are called "co-hearers." Now we turn our attention to hearers who are not directly participating. The speaker may tolerate or attempt to exclude any not direct participants in the conversation. The tolerated hearers are termed "over-hearers" and those excluded deliberately from the utterance are termed "eavesdroppers". For the purposes of this thesis we exclude all not direct participants of the utterance.

Accessibility to an utterance	
Ratified Participants **Direct Participation of the recipient to the utterance**	**Unratified Participants** **not direct participation of the recipient to the utterance**

Addressed By the speaker	**Unaddressed** By the speaker	**Tolerated** By the speaker	**Excluded** By the speaker
Conversation Partner	Co-Hearer	Overhearer	Eavesdropper

Table 2 *A participation framework model for hearers. Adapted from Goffman (1981).*

We now undertake a more specific look at this framework to analyze how a hearer progresses from one state to another in our system. In an educational environment, it is fair to assume that a lecturer intends for the entire class to hear and directly participate in the discourse. This is not to say that a lecturer may not intend to address one particular learner to say answer a question and solicit follow up, but the model the author posits assumes all participants in the exchange are ratified participants. These participants may take the space of a conversational partner, a co-hearer, or even a speaker. Movement through these states is shown in **Figure 1**.

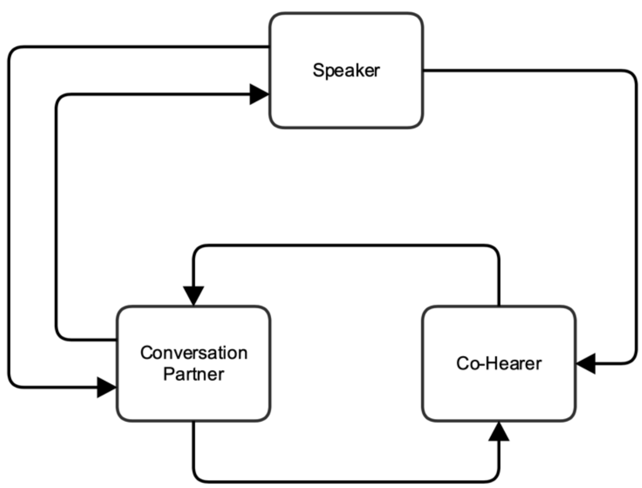

Figure 1 *A model of the participation framework for hearers showing the various states an individual learner can occupy.*

Now we explore the role of the speaker producing the utterance. Each utterance is made from two components:

- the syntactic structure, the word choice and formulation of the utterance
- the semantic contribution, the content of the utterance

A speaker can be responsible for one, both, or neither of these two components of an utterance. Speakers who are responsible for both the syntactic and semantic components of his or her utterance are called the "authors" of the utterance. Speakers who are not responsible for either components are termed "relayers" of the utterances. An example would be when Alice broadcasts a job advertisement composed by Bob. A "ghostee" is a speaker is responsible for content of the message but not the devising the wording and structure of a message. An example is where Alice asks Bob for a form of words to express a certain message which she wanted to send. Lastly, we have the opposite scenario where a speaker is responsible for the formulation of an utterance but takes the content of a previous utterance. We call these individuals are termed "spokesman" (Levinson, 1998).

	Responsible for the **Content** of an Utterance	Responsible for the **Formulation** of an Utterance
Author	+	+
Relayer	-	-
Ghostee	+	-
Spokesman	-	+

Table 3 *A participation framework model for speakers. Adapted from Levinson (1998)*

As a learner progresses through the topic we expect to see the learner progress form a realyer to an author when expanding questions to the instructor as seen in **Figure 2**.

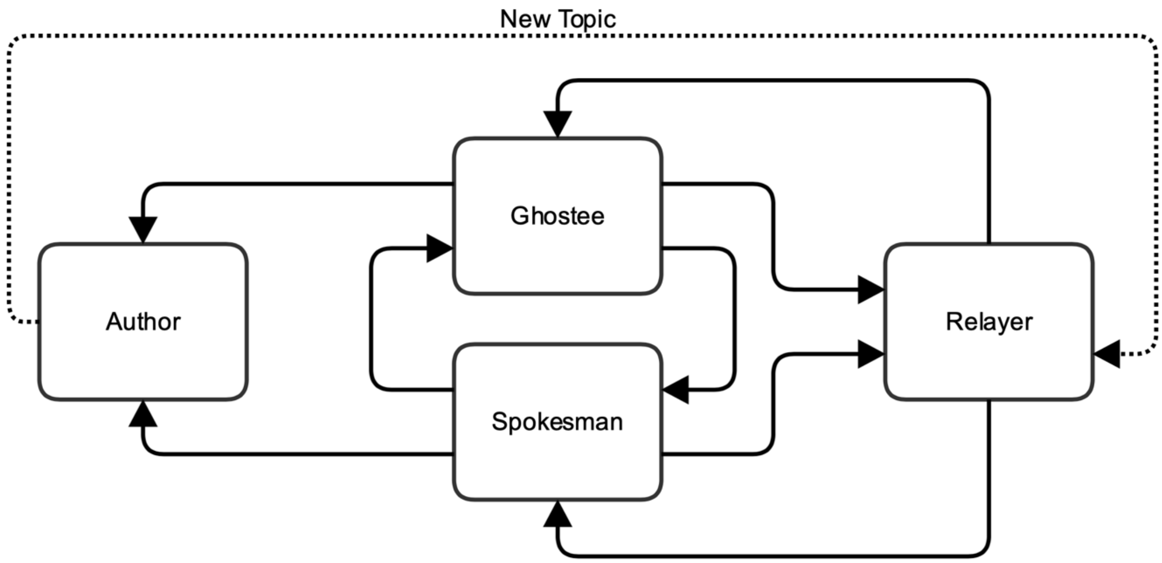

Figure 2 A model state diagram for the participation framework for speakers

4.2. Theoretical Curriculum Evaluation & Development Model

This thesis utilizes a variant on the curriculum development oriented to improving learning competences model (Felder & Brent, 2003; Gonzalez & Wagenaar, 2003; Mendez et al., 2014; Tobon, 2007). This process generally follows this systematic approach: precise identification of the curricular design specifications and constraints (e.g. student outcomes, competences, and learning goals) informed by stakeholders and society; developing and testing/evaluation of the curricular design; and refining the design with the feedback of students and stakeholders (Felder & Brent, 2003; Mendez et al., 2014). This process is iterative and assessment is an important component of feedback from the entire curriculum development process, generating opportunities to improve it. **Figure 3** below, depicts the typical learning-outcome-centered curriculum based on the systematic approach described above incorporating the assessment generated by the TRACKR engagement model.

al., 2009):

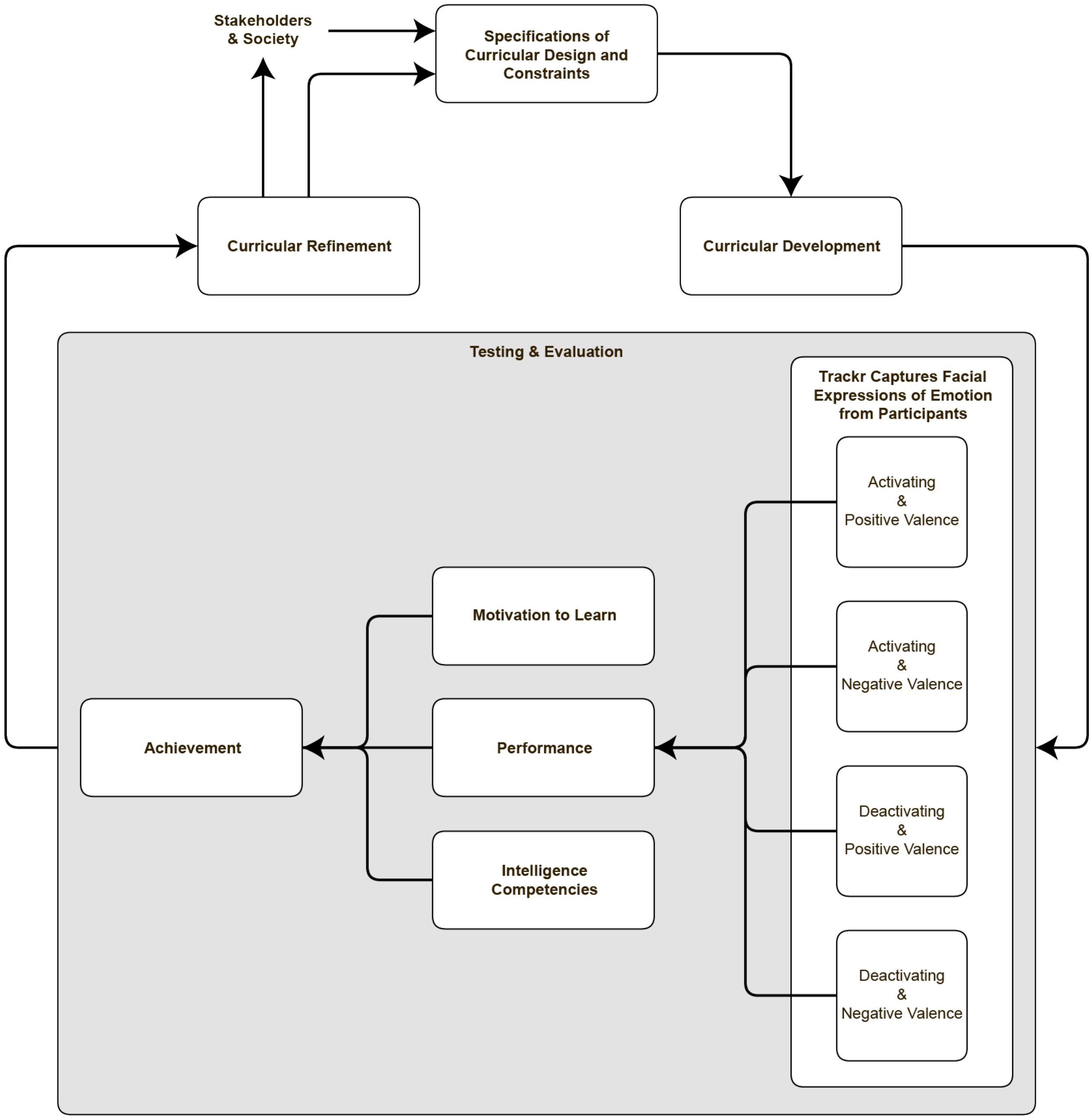

Figure 3 *TRACKR Framework and Model*

The testing and evaluative section of the model attempts to find relationships between traditional assessment, measures of knowledge transfer, and FEE. This model is theorized as a simple summation of the respective emotions with normalized proportionality coefficients (Pekrun et al., 2009):

$$f_{performance} = a_{act}\left(k_1 e_{act,pos} + k_2 e_{act,neg}\right) - a_{deact}\left(k_3 e_{deact,pos} + k_4 e_{deact,neg}\right)$$

$$\forall\ k_1, k_2, k_3, k_4\{-1,1\}$$

Equation 1 *The principal performance measure utilized by the TRACKR Framework*

This model, shown in **Figure 3** above, is also used continuously as part of the larger research model, as a type of participatory action research model which allows the research team to reflect on the curriculum effectiveness and present preliminary findings to the instructors participating in the thesis.

5. Methodology

In this section, the experimental setup to acquire the test dataset, methods of data analysis and their correspondence to learner behavior is presented. The goal of the experiment was to create a data capture system to test the aforementioned model.

5.1. Procedure

Participants included fifteen graduate and undergraduate students from a private university in the United States. The experimental dataset was obtained during four class sessions in a stadium set up classroom. Students were asked to follow a lecture and take notes and answer questions during the sessions. A video of the 90-minute lecture sessions was recorded by three high definition (*HD*) cameras in each of the lectures. Two *HD* resolution (*1920* by *1080* pixels) cameras were located at the front of the classroom above the chalkboards on the left and right sides of the room aimed at the students to record FEE and body pose. A third *HD* resolution camera was located at the rear of the classroom aimed at the lecturer to provide additional context to facilitate interpretation of the results. A system was put in place to record the lecturer’s digitally presented content on the in-room projection screens. This system was designed to synchronize the lecture content with the content of the student's FEE to facilitate analysis between multiple views of the same environment, see **Figure 4** below. In summary, this setup maintains four synchronized video feeds captured at high definition content for automated offline parsing.

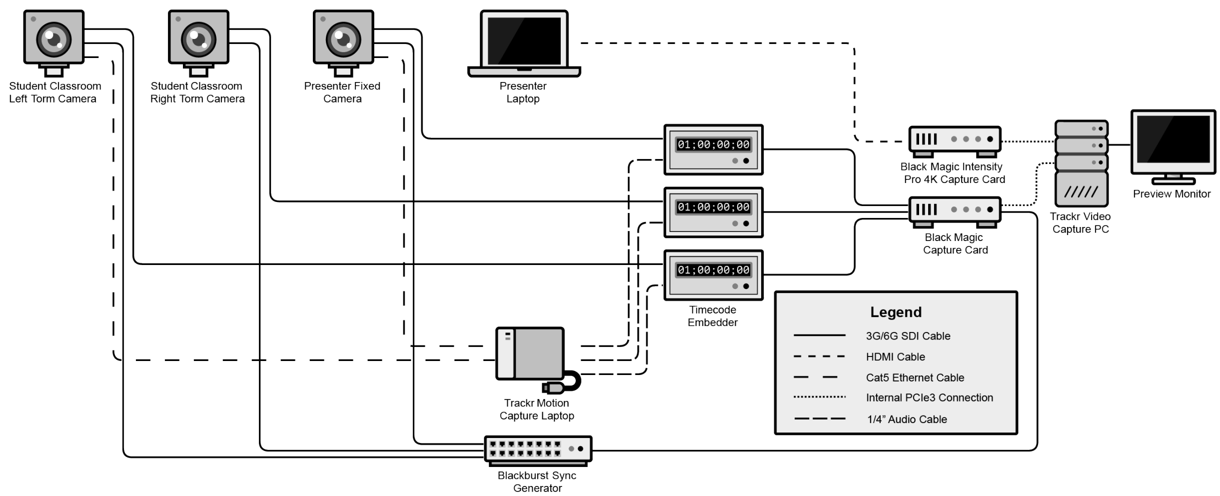

Figure 4 *A System Architecture Overview for* TRACKR

5.2. Measures

The recording system extracted and stored several types of data: a) Color frames from the HD resolution cameras extracted at one frame per second; and b) Images. These images that resulted from these frames were then processed using a multi-view face detection to isolate the faces for further processing. The most recent and state-of-the-art approach was adopted because it is the most accurate and the fastest (Chen et al., 2014). Notably, it is significantly faster than the previous best methods in terms of computation time and memory. Such high-speed low memory systems are crucial for a real time face detector as well, which will be critical for future research.

5.3. Processing

The offline processing and analysis of the extracted data was performed by *Node.JS* scripts. The resulting isolated faces were then analyzed to label the FEEs. The system labels FEE for the following positive valence emotions: happy, surprised, and calm; and negative valence emotions: anger, sad, confusion, disgust, and fear. The emotions that appear to be expressed on the faces provide a confidence level as a value bounded between zero and one. Each face can demonstrate more than one of the aforementioned FEEs.

The system then tracked all labeled FEE for each of the participants throughout the session. At each time step available the system takes each individual participant's labeled FEE and calculates their performance score according to **Equation 1** shown above. The resulting scores were then averaged among all participants during that time step to ensure the anonymity of each participant, while still providing generalizable measures of knowledge transfer to the participants.

The prime advantage of using synchronized camera and digital lectures to study engagement is that they can provide detailed and descriptive accounts of the contextual factors taking place when different FEEs took place, providing us with contextual factors of students higher or lower affective engagement levels. These descriptions enhance our understanding of unfolding processes within contexts. The major advantage of these techniques is that it reduces the labor-intensive process of regular observations, and avoid the limitations of only being able to observe a small number of students and contexts. Further, the quality of descriptive observations increases as it will no longer depend on the skills of the observer and on his or her ability to capture and make sense of what was observed.

In order to make sense of the average class session performance scores based on the system labeled FEE, the scores were cross referenced with the synchronized camera and digital lecture video feeds to verify that the trend of the scores matched the contextual factors in the learning environment (e.g. a participant asking a question, the lecturer showing a video demonstration, etc.).

6. Analysis

In total four class sessions of video were recorded for analysis by the TRACKR framework of the fifteen participants in a stadium set-up classroom. The video was recorded using remotely operated pivot-tilt-zoom (PTZ) cameras. The classroom used was already equipped with HD Sony BRC-H900 PTZ Cameras. After a process of manually cleaning up the resulting videos, ensuring that participants were in camera focus, visible and not overlapping in their primary seated position, as well as ensuring coordinated synchronization between the four video feeds one approximately 90-minute session of video was obtained.

The 90 minute video can be broken up into a 15 minute “pre-roll” session of participants trickling into the classroom before the stated start time of the experimental session, 6 minute session where the lecturer waits for participants who are arriving late often termed the “[Redacted Private University] standard grace period” by the lecturer, a brief 10 minute session where the lecturer introduces a class room activity to be conducted in small groups, the students participate in this group activity for the subsequent 25 minutes, and then for the remaining 12 minutes the lecturer presents a video segment covering similar topics with added commentary. The last approximately 5 minutes of video are students hanging back to ask questions to the lecturer and teaching assistants in an ad-hoc manner.

The video was trimmed to a 59-minute section beginning with the lecturer introducing the classroom activity and ending with the lecturer dismissing the participants. This session was broken down into 59 frames of video content of the fifteen participants. Due to the participants and lecturer moving around the classroom during the group work section this provides a total of 244 individually isolated faces that were further processed by the TRACKR framework into 1952 confidence predictions for labeled FEE.

The predictions for each of the labeled FEE among all the faces detected during a given time step were averaged together to yield a classroom-wide measure of that FEE. To better understand the impact of each measured "classroom FEE" we take a look at the distribution throughout the recorded session to better understand the baseline measurements. The classroom FEE (n=244) averaged 32.62+10.63 through the recorded session. Further, the measurements for each of the labeled FEE are broken apart in **Figure 5** below. Most of the classroom FEE appear to be distributed equally, though the average prediction for the positive valence emotions happy and calm appear to be different.

The overall distribution of labeled FEE does not come from a normal distribution, as measured by a combined omnibus test of normality (D'Agostino & Pearson, 1973; Oliphant, 2007). Therefore, a two-tailed Mann-Whitney U nonparametric test with H_0: "the distributions of all labeled FEE and the measured FEE populations are equal" and H_A: "the distributions are not equal" with $\alpha = 0.05$ was conducted (Mann & Whitney, 1947; Oliphant, 2007). This test yields that the distributions for the positive valence emotions calm and the negative valence emotions: sad, disgust, and fear reject H_0 indicating that there is a statistically significant difference between these classroom FEE and all the other FEE.

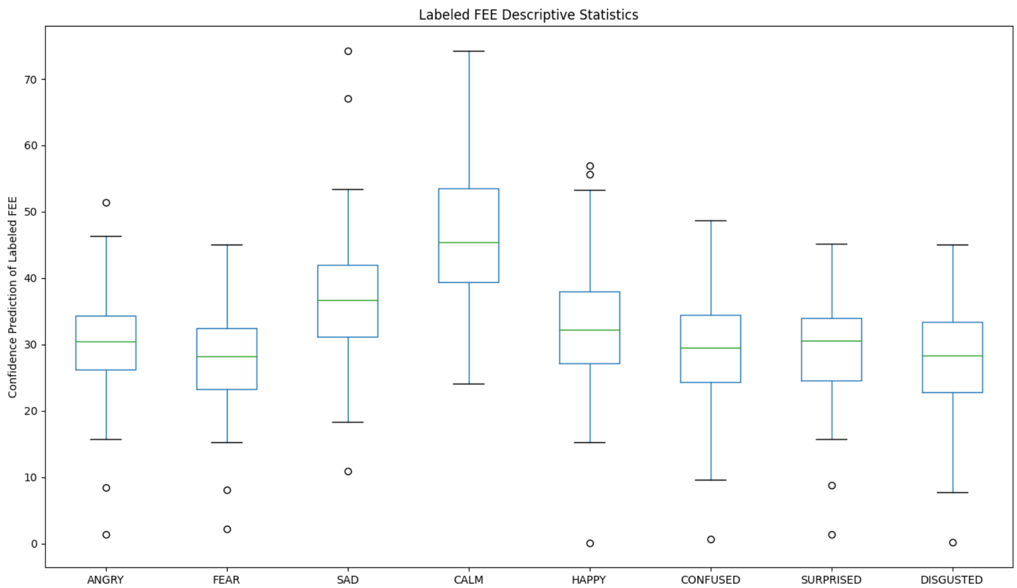

***Figure 5** Labeled FEE descriptive statistics throughout the recorded session.*

The data exploration continues through plotting the classroom FEE against each time step as seen in **Figure 6** below. In this plot the positive valence emotions surprised and negative valence emotions: anger, disgust, and fear tend to overlap and follow each other.

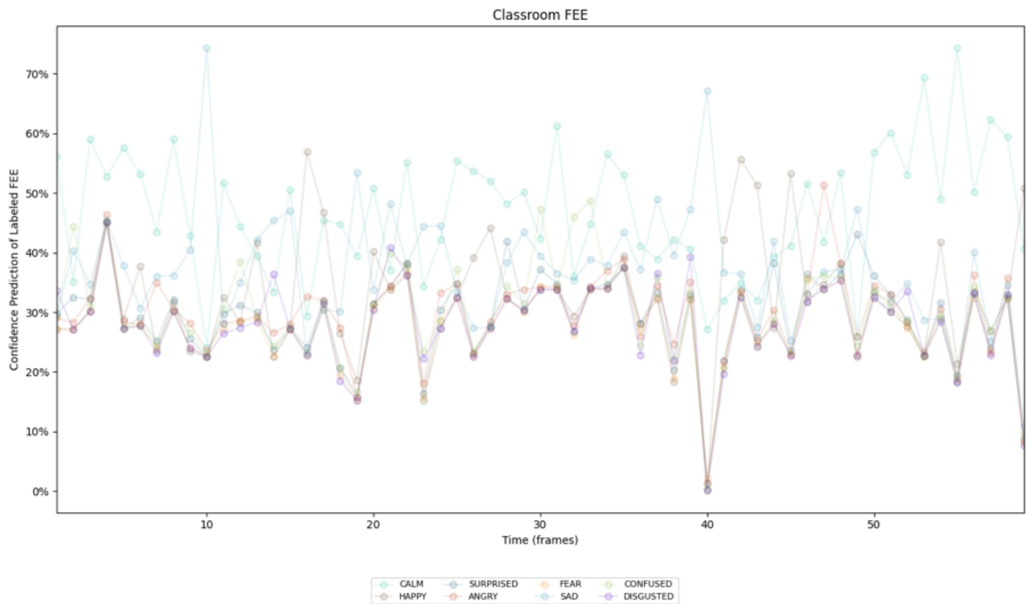

Figure 6 A plot of the classroom FEE (comprised of the average predictions for a given FEE for all participants during that time step) during the recorded session.

To further explore the relationship between these emotions, they are extracted and plotted on their own chart see in **Figure 7** below. Additionally, a robust locally weighted regression to smooth the scatterplot is plotted, as the classroom FEE have dramatic changes in prediction from one time step to the next (Cleveland, 1979; Seabold & Perktold, 2010). Furthermore, the plot includes a 95% confidence interval on the measurements provided to facilitate an easier comparison.

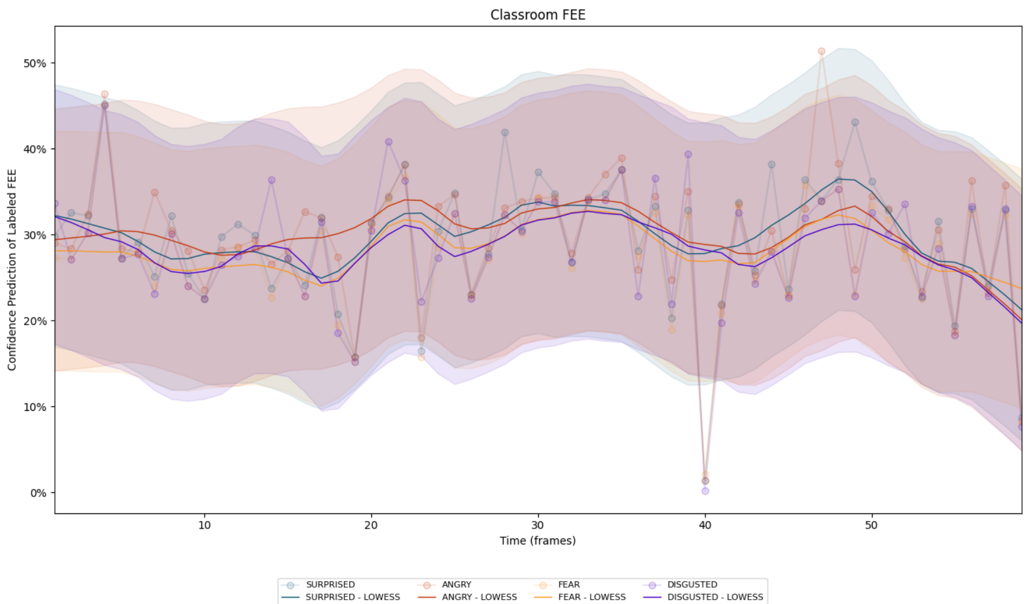

Figure 7 *A plot showing the positive valence emotions surprised and negative valence emotions: anger, disgust, and fear with a robust locally weighted regression fitted with a 95% confidence interval on the regression parameters.*

From **Figure 7** the author concludes that these classroom FEE belong to the same distribution. This is confirmed by performing another set of two-tailed Mann-Whitney U nonparametric tests with each of the positive valence emotions surprised and negative valence emotions: anger, disgust, and fear against each other and the tests failed to reject ($\alpha = 0.05$) the null hypothesis, confirming that they are indeed from the same distribution. As such for future reference in this work, the author will take the average of the classroom FEE for negative valence emotions: anger, disgust, and fear together and report them as "combined negative valence" FEE. An updated plot showing this combined negative valence FEE alongside the other FEE is shown in **Figure 8** below.

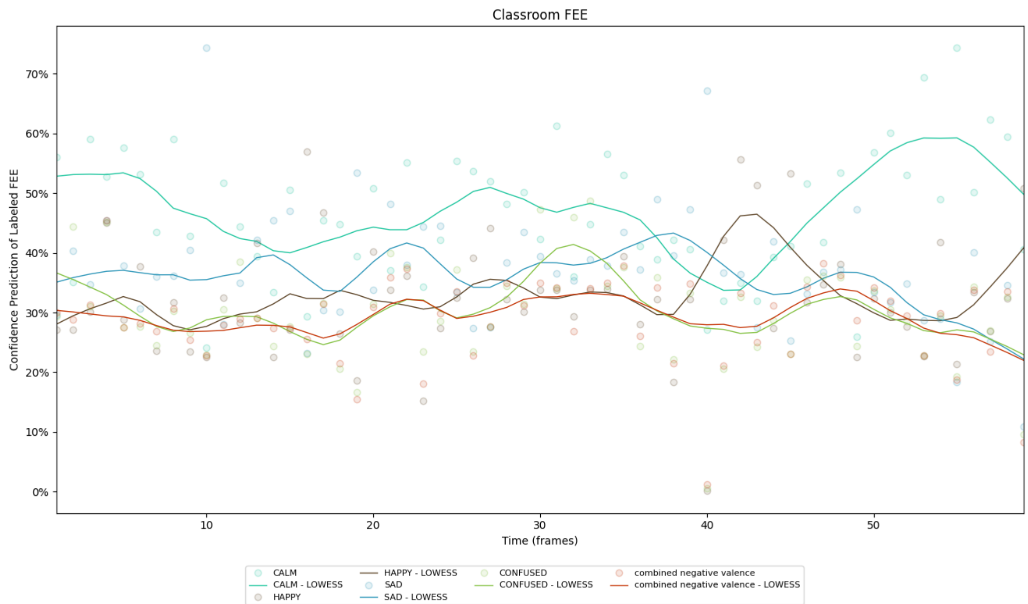

Figure 8 *A plot of the classroom FEE with the new combined negative valence FEE—the average of the negative valence emotions: anger, disgust, and fear—during the recorded session.*

This new plot, **Figure 8**, provides additional clarity to understand the relationship between the classroom FEEs over the recorded session.

Next, the classroom FEEs are centered and scaled to unit variance, to begin to explore the normalized proportionality coefficients as described in **Equation 1**, above. Now, with a centered and scaled and independent set of distributions for all of the classroom FEE, the positive valence emotions: happy and calm are transformed. The classroom FEE for surprised was removed from the set of positive valence emotions as it belongs to the same distribution as the combined negative valence FEE and is therefore accounted in the model already. The analysis was conducted with the method of principal component analysis of the two FEE in order to reduce the dimensionality of the FEE while maximizing the variance in a linear fashion as theorized in the TRACKR framework (Halko et al., 2009; Jolliffe & Cadima, 2016; Seabold & Perktold, 2010).

This yields a new measure for the positive valance emotions, which I've demarked as "PCA-PVE" which can be described as follows, in **Equation 2** below:

$$f_{PCA,PVE} = k_{happy}e_{happy} + k_{calm}e_{calm}$$

$$\approx 0.707\ e_{happy} - 0.707\ e_{calm}$$

Equation 2 *Principal component analysis for the positive valance emotions: happy and calm to yield a newly reduced FEE*

The results of this data transformation can be seen in **Figure 9**, below, where the classroom FEE for the positive valance emotions and the resulting transformation are all plotted on the same axis with the corresponding locally weighted regression.

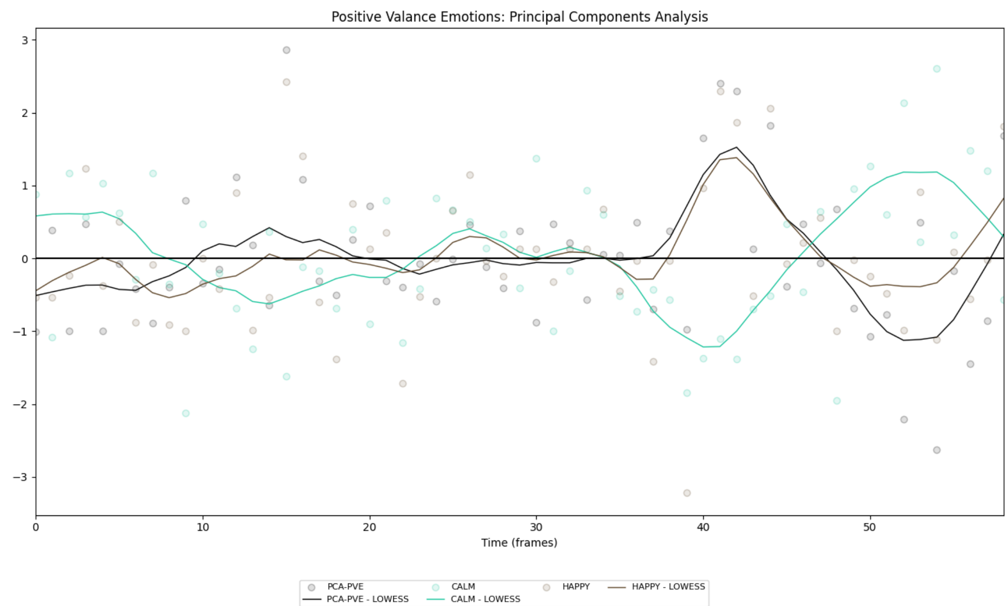

Figure 9 *A plot of the positive valance emotions classroom FEE as well as the newly reduced PCA component during the recorded session. We can see that the transformation represents the data well*

It is evident that the transformed data seems to follow the trends of the underlying classroom FEE, and appears to be a good measure for the positive valance emotions.

A similar process was taken for the negative valance emotions: sad, confused, anger, disgust, and fear. This yields a new measure for the negative valance emotions, which I've demarked as "PCA-NVE" which can be described as follows, in **Equation 3**:

$$f_{PCA,NVE} = k_{sad}e_{sad} + k_{confused}e_{confused} + \frac{k_{cnv}}{3}\left(e_{angry} + e_{fear} + e_{disgusted}\right)$$

$$\approx -0.025e_{happy} - 0.707\, e_{calm} - 0.236\left(e_{angry} + e_{fear} + e_{disgusted}\right)$$

Equation 3 *Principal component analysis for the negative valance emotions: sad, confused, angry, fear, and disgusted to yield a newly reduced FEE.*

This means that 'preserving as much variability as possible' translates into finding new variables that are linear functions of those in the original dataset, that successively maximize variance and that are uncorrelated with each other.

Similarly, again the results of this linear transformation can be seen in **Figure 10**, below, where the classroom FEE for the negative valance emotions and the resulting transformation are all plotted on the same axis with the corresponding locally weighted regression. In this transformation, the maximizing variance procedure has inverted the relationships between the original curves and the linearly transformed data; however, this poses no concern as it would just necessitate an inverted constant later.

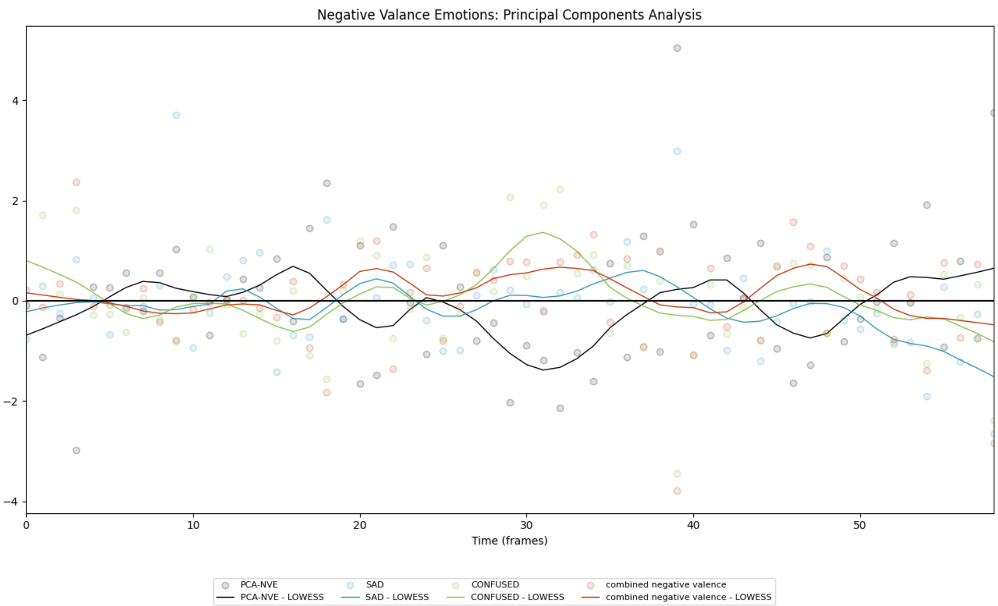

Figure 10 *A plot of the negative valance emotions classroom FEE as well as the newly reduced PCA component during the recorded session. From the plot, one can see that the transformation has inverted the relationship but this will be accounted for by the multiplicative constant in the final relationship.*

Lastly, with these two reduced and transformed components representing both the positive and negative valance emotions, **Equation 1** can be updated as follows:

$$f_{performance} = a_{act}\left(k_{happy}e_{happy} + k_{calm}e_{calm}\right) - a_{deact}\left(k_{sad}e_{sad} + k_{confused}e_{confused} + \frac{k_{cnv}}{3}\left(e_{angry} + e_{fear} + e_{disgusted}\right)\right)$$

Equation 4 *An update to measured performance indicators with the linear transformation from the PCA.*

This simplified equation allows us to compose a component score for both positive and negative valence emotions, this simplified data transformation is again plotted on the same axis with the corresponding locally weighted regression, seen in **Figure 11** below.

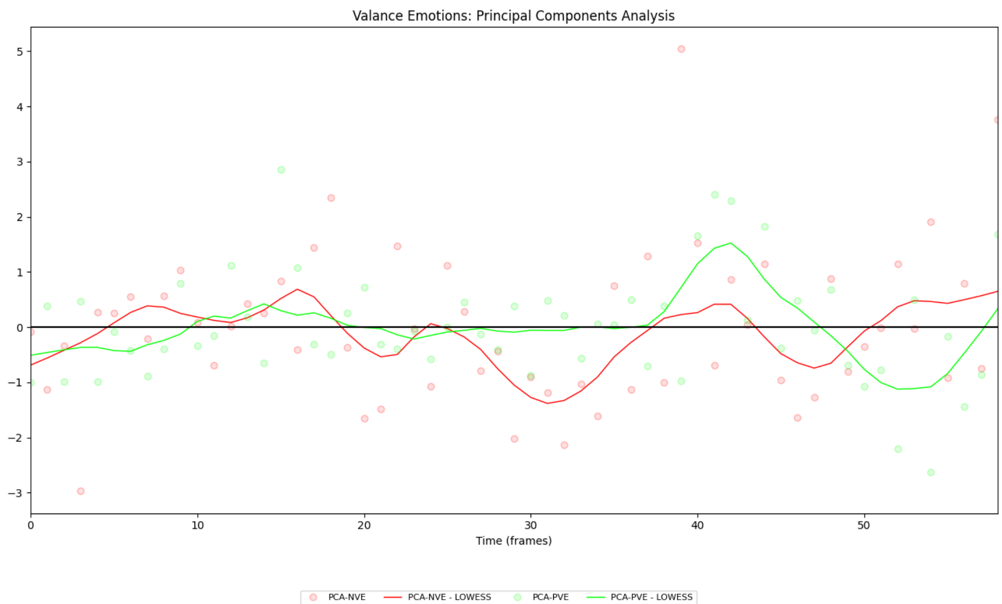

Figure 11 *A plot of the valence emotion reduced PCA components during the recorded session.*

It seems that the two PCA reduced classroom FEEs have a periodic underlying structure with major "events", or crests in the in the regression curves, at the 7, 16, 24, 41, and 53 minute marks. These *events* correspond with the overall structure of the class environment as shown when you take these events and overlay them on the description of the recorded session, presented earlier, as seen in **Table 4** below. This seems to indicate that there is a relationship between the classroom FEEs and the structure of the learning experience; however, absent any other measure, it is difficult to ascertain the relationship between the classroom FEEs and a participant's knowledge transfer.

Start	End	Description
00:00	*00:06*	Participants arriving late, lecturer providing “[Redacted Private University] standard grace period”
00:07	*00:16*	Lecturer introduces a small group activity
00:17	*00:24*	Participants participating in the small group activity
00:25	*00:41*	Participants participating in the small group activity
00:42	*00:53*	Lecturer presents a video segment covering similar topics
00:54	*00:59*	Lecturer summarizes the points of the video

Table 4 *A breakdown of the major "events" shown from the two PCA reduced classroom FEE and the recorded session agenda.*

7. Discussion

The two PCA reduced classroom FEEs of the positive and negative valance emotions are able to capture and represent the collective learning experience of all participants in a recorded session. It appears that there is also a relationship between these classroom FEEs and the structure of the learning experience. Moreover, this captured experience can only be fully realized through the collection and analysis of both positive and negative valence emotions in classroom FEEs. Only when taken together there is enough resolution in the transformed data to extract higher level meaning. These results are in keeping with the theoretical work on AEs furthered by the CVT model, that is to say, the nature of an emotion by itself is neither useful or harmful, rather, it must be considered within its context specific to a learning activity or an outcome (Pekrun, 2006).

AEs, as conceptualized by the CVT model, are amplified by and benefit from a temporal understanding. This allows for variation in AEs learning up to, during, and after an assessment activity or learning outcome; however, the popular AEQ measure is a student self-report survey measure, which due to experimental constraints are measured at discrete intervals and cannot capture a student's AEs throughout the entirety of an assessment activity. In contrast this aforementioned new automated camera vision technique is able to measure emotional states using machine learning algorithms at a resolution limited only by the camera equipment and computing power allowing for more complex temporal understandings of AEs.

The proposed TRACKR system conducts a classroom wide search of AEs taking a snapshot of all participants FEEs to derive the state of the classroom and all its participants. Traditional observational measures, on the other hand, typically involve a small number of students as a representative sample of the whole. Through measurements of the entire classroom, the TRACKR system ensures that the struggle or ease of one individual participant with the given material is not generalized to the entire cohort or even larger settings. This improvement reduces the labor-intensive process of traditional observations methods, which are costly to administer (Fredricks & McColskey, 2012). Additionally, the two PCA reduced classroom FEEs are also more robust to outliers due to their composition as a linear transformation of multiple averages, as seen in **Equation 4**, which ensures not only that any insights seen in the data are caught by one or more of the various measured FEEs, but also that these measures are not individually identifiable, protecting the privacy of the participant.

The insights gained from the two PCA reduced classroom FEEs representing positive and negative valance emotions can be used as substitute qualitative measures to replace the popular AEQ measure as a "more sophisticated measures of emotions and their components … [and] methods for analyzing [their] multivariate functional relationships over time," an area indicated for further research by Pekrun in "The Control-Value Theory of Achievement Emotions" (Pekrun, 2006). In the CVT model there are multiple implications of AEs for educational practice; however, due to the nature of the classroom-wide measure employed in the TRACKR system not all of the recommendations in the literature apply, as they are focused on individualistic interventions.

The principal factor influencing the participants' control, values, and emotions is likely the quality of instruction, which pursuant to CVT increases the participant's sense of control and his/her positive academic values. These changes would be reflected in new valuations by the participant and observed in his/her FEEs. This intervention would improve aspects of instruction to increase the clarity, structure, and presentation of tasks. This could also be updated by informing the relative match between participants' capabilities and the materials needed to be mastered. Under the CVT model, if demands are too high or too low, participants may experience boredom or anxiety (Pekrun, 2006).

This intervention to increase the participants' control mentioned in the CVT model is more thoroughly fleshed out in the curriculum development oriented to improving learning competences model (Felder & Brent, 2003; Gonzalez & Wagenaar, 2003; Mendez et al., 2014; Tobon, 2007). In the iterative TRACKR model the performance measures generated by the two PCA reduced classroom FEEs can be used to ascertain the instructional contexts that need reemphasis, teaching or learning methods that need to be revised and resources that need to be rewritten. Specifically, curriculum developers should look towards balancing any "spikes" or large changes in the two measures as they indicate a mismatch between that event and the larger learning outcomes. This development process is illustrated in **Figure 3** on page 20 above.

8. Limitations of Research

This thesis presents labeled FEEs as a series of expressions through a given class session as a function of time; however, there would be additional benefit gleaned by incorporating a rigorous coding of instruction activities to better contextualize the FEEs. The theoretical framework posits that different instructional activities will induce different AE's that should be interpreted contextually in order to glean generalizable information from the participants knowledge transfer.

The camera system utilized by this thesis encountered difficulties with occlusion and obstruction during classroom group work activities where participants re-arranged themselves into small groups. This re-arrangement partially occluded participants from one or more camera angles making it difficult to isolate the faces with the machine learning algorithms for further processing. These temporary obstructions make it difficult to obtain a reading on all participants which can unintentionally impact everything from the raw data to the interpretable results.

Furthermore, this thesis aimed to unobtrusively monitor participants during a real-life learning session with no built-in assessments for measuring mastery of the material covered. This restriction made it difficult to draw any conclusions about the actual knowledge transfer during a class session. Instead this thesis focuses on drawing parallels between the two PCA reduced classroom FEE measures and the classroom behavior much the way other measures utilize instructor report on students (Birch & Ladd, 1997; Buhs & Ladd, 2001; J. D. Finn et al., 1991, 1995; Skinner & Belmont, 1993; Wigfield et al., 2008). This observational method shown in the prototype of the TRACKR system above, similarly struggles from the abilities of an observational system to properly interpret the affective engagement of the student. Unlike these methods, due to the video records produced, the measures employed in the TRACKR system are able to assess all participants in a given classroom.

One of the difficulties in labeling the positive and negative valence emotions were the differences in terminology between the machine learning algorithms used and the theoretical constructs utilized. These differences reduced the available labeled FEE that could be used to asses affective engagement among participants. These effects of these restrictions are difficult to quantify; however, the CVT model describes sixteen emotions, as shown in **Table 1** on page 12, as opposed to the machine learning model employed which provided eight labels, only half of those employed by CVT (Chen et al., 2014; Pekrun et al., 2007).

The unobtrusive nature of the prototype created, did not explore qualitative measures or interviews to better understand why participants in the learning sessions expressed a given FEE. These intraindividual, occurring within an individual participant, psychological functions can be used to predict and explain individual differences that are represented in the larger classroom FEEs (Pekrun, 2006). Incorporating these techniques in future research will create opportunities for further refinement of the theoretical construction.

This thesis relies on one machine learning algorithm in order to ascertained the likely labeled FEEs; however, recent studies have shown that machine learning algorithms can discriminate based on race and gender (Buolamwini & Gebru, 2018). Thus, the aforementioned results are limited in this respect as the research does not account for these differences, though the population of the sample analyzed does have various participants who are more likely to be misidentified and consequently have mislabeled FEEs. This could be corrected by using an ensemble system of machine learning that incorporates a variety of predicted labeled FEEs, as these algorithms use different methods to classify the FEEs based on different training datasets. This would replicate the behaviors visible in multi-observer classroom engagement measures as outlined previously (Fredricks & McColskey, 2012).

Lastly, the research above measured the labeled FEEs of participants while lectured by one specific lecturer. This could pose a difficulty for understanding the deeper relationships between a given learner's knowledge transfer as it conflates the material being disseminated and the individualized delivery of the lecturer.

9. Suggestions for Future Research

The TRACKR framework presented above demonstrates the potential power that an automated system can provide to validate the broader cumulative knowledge for emotions and participants' learning and achievement.

Academic educational researchers specializing in achievement emotions have remarked that strategies of analyzing participants emotions benefit from a combined approach of both qualitative and quantitative strategies. Qualitative exploratory analysis lay the ground work for constructing measures and undertaking quantitative studies (Pekrun et al., 2002). Quantitative studies are able to test the generalizability of findings to ensure that conclusions are valid for individual participants. In a later study, some authors propose the value of integrating observational systems for facial and postural emotion expressions into video-based classroom observation to analyze participants' and lecturers' ongoing emotions in a classroom. The proposed TRACKR system posits a similar improvement (Pekrun, 2006).

One of the most pressing areas for future research and improvement is a harmonization between the theoretical construction of emotions within educational and psychological research more generally and the coding schemes employed taken as ground truth for machine learning algorithms within computer science research. In particular, the manual annotation systems to evaluate and measure behaviors, emotions, or perceptions are often designed to allow for novice annotators to perform annotations of data displaying spontaneous FEEs (Dupre et al., 2015). Yet educational researchers theoretical frameworks are based on a trained skilled observer (Fredricks & McColskey, 2012; Turner & Meyer, 2000). This harmonization effort would require significant genuine dataset development in order to catalogue a large representative sample of learners in multiple classroom environments on which to train, validate, and test an existing state of the art machine learning algorithm. This updated model can be tuned to perform as good as a human trained observer.

Next, a laboratory study should be used to assess the basic mechanisms of mood and human emotions, within artificial constraints to show potential causality. A simulated training where new entry-level construction workers would be presented with off-site safety trainings to learn a particular safety-related task that they have no experience with, or are unlikely to have experience with in the next six months would be established. This training should be broken up into distinct sections covering one topic; knowledge assessments should be given after each section to serve as a baseline for the effects of knowledge transfer. Additionally, immediately after the entire training, participants would take a knowledge retention test to gauge understanding of the safety knowledge and a risk judgement test to understand safety motivation on the trained material. Participant's personality and risk aversion may confound results by making them more aware of their surroundings of safety risks or more likely to judge a behavior as risky independent of the safety training, to control for these traits. Participants will also take a conscientiousness test using the *NEO-FFI* subscale (McCrae & Costa, 2004), and a domain specific risk perception using the *DOSPERT* scale, respectively (Blais & Weber, 2006). To measure long-term knowledge transfer, another test would be issued with a follow-up test 4 months later. While the duration of time any construction worker spends on a given site depends on a multitude of factors, with some individuals coming in for a few days and others staying on site year after year, studies have shown that the average time on site is 0.93 months and the median length of time is 1 month. Furthermore, the "half-life" of the on-site workers is approximately one month, with less than 5% of workers staying on the same site four months after starting (Sparer, 2015). By re-testing knowledge transfer after 4 months, the system measures the long-term effects relative to the majority of workers time on site. The training session should be recorded in keeping with the set up described earlier in section 5 Methodology above. Using these scores on a participant's knowledge transfer as inputs to train the parameters of the TRACKR system to learn a relationship between the behavioral and affective engagement cues from the camera vision system inputs and a participant who has been able to transfer the knowledge from the training.

In order to inform education practice and occupational health and safety in validated ways in the near future, a field study must also be undertaken. Specifically, field experiments can monitor the effectiveness of educational interventions within real-life, context-bound, intense AEs experienced by participants in educational environments, which would be difficult if not impossible to replicated inside a laboratory due to ethical constraints.

As such, in order to validate the effects of the CVT model and the TRACKR system on adult learners requires partnering with at least two "real-life" training companies to measure participants reactions in an appropriate context. To control for management practices and turnover each firm would provide two comparable training sites, matched in scope of work, years of experience for trades, union/non-union shops, subcontractor safety records, highest laborer educational attainment level, shared senior management, geographic region, and corporate culture to serve as pre-existing clusters. One of these sites at each firm, chosen at random, would serve as a control group with the traditional training paradigm employed by that firm at those sites. The other site, will employ the TRACKR variant on the curriculum development oriented to improving learning competences models. At this site, unobtrusive cameras will be installed in the off-site training facilities to ensure sufficient coverage to obtain views of each participants FEEs, at a resolution compatible with the newly harmonized machine learning algorithm. The cameras shall have a shutter speed and automatic focus adequate enough to obtain clear imagery for the TRACKR prototype system. Additionally, there shall be additional cameras in order to capture the lecturer, and any other contextual factors (e.g. presentation screens and materials, live demonstrations, etc.) relevant to the training environment. These cameras shall be frame synchronized to ensure that each frame across all cameras represents the same snapshot moment in time to coordinate the classroom FEE for the environment. The feedback generated from the trained parameters of the TRACKR system on behavioral and affective engagement cues from the camera vision system input will be used to estimate if participants will be able to transfer the knowledge from the off-site training to the field. These estimates will be provided to trainers to modify the training curriculum to cover topics more in depth, review material, or move quickly past it. Specifically, trainers will look towards balancing any “spikes” or large changes in the two PCA reduced classroom FEE measures to avoid mismatches between that event and the larger learning session. Given that the curriculum development oriented to improving learning competences model is iterative these changes from one group of participants can be taken to improve the standard curriculum for the next group. This necessarily implies that the subsequent groups will benefit from the structural changes to the instructional contexts that needed reemphasis, revised learning methods, and rewritten resources. This iterative process should be repeated as often as necessary as indicated by the feedback from the TRACKR system.

Similar to the laboratory environment, participant's personality and risk aversion may confound results, therefore to control for these traits, participants in both the control and treatment groups will also take a conscientiousness test using the *NEO-FFI* subscale (McCrae & Costa, 2004), and a domain specific risk perception using the *DOSPERT* scale, respectively (Blais & Weber, 2006). Additionally, in order to understand the success of knowledge transfer between the trainers and the participants, all participants shall be provided with a pre and post-tests on the material covered.

Lastly, to validate the relationship between more engaged participants during a training session and improved safety outcomes on site, all site safety records, including warnings, infractions, accidents, and praise from when the participants in the study first step foot on site until the last has left shall be qualitatively observed for patterns indicating successful knowledge transfer or lack thereof. This information can also be qualitatively measured using a weighted scoring system based on the severity and occurrence of the underlying report (Sparer, 2015).

To date, there are few available successful intervention studies on emotions, but there exist frameworks within affective analysis of text anxiety research that suggest it is possible. This work is not an easy task, but if implemented can prove techniques to inform occupational health and safety training in validated ways in years to come.

10. *Conclusion*

This thesis proposes a modified curriculum development model oriented to learning *transfer* as an iterative method to improve safety outcomes. This above proposed framework measures construction workers' engagement during construction training courses conducted by unobtrusively analyzing engagement through body and pose estimation, codifying who is speaking, and understating the predicted emotional state of a given worker through their facial expressions of emotion through state-of-the-art computer vision techniques. The subsequently transformed classroom average indicators of the positive and negative valance emotions are able to capture and represent the collective learning experience of all participants in a recorded session. These indicators can be used to ascertain the instructional contexts that need reemphasis, teaching or learning methods that need to be revised, and resources that need to be rewritten. ¤

11. References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Birch, S. H., & Ladd, G. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, *35*(1), 61–79.

Blais, A.-R., & Weber, E. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*, 33–47. https://doi.org/10.1037/t13084-000

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology*, *37*(4), 550–560.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Bureau of Labor Statistics. (2017). *National Census of Fatal Occupational Injuries in 2017 [Press release]*. 9.

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative Effectiveness of Worker Safety and Health Training Methods. *American Journal of Public Health*, *96*(2), 315–324. https://doi.org/10.2105/AJPH.2004.059840

Byrnes, J. P. (2007). *Cognitive Development and Learning in Instructional Contexts* (3rd ed.). Allyn and Bacon.

Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J. (2014). Joint Cascade Face Detection and Alignment. *ECCV*. https://doi.org/10.1007/978-3-319-10599-4_8

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *The Journal of Applied Psychology*, *94*(5), 1103–1127. https://doi.org/10.1037/a0016172

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cohen, A., & Cooligan, M. J. (1998). *Assessing occupational safety and health training—A literature review.* https://doi.org/10.26616/NIOSHPUB98145

D’Agostino, R., & Pearson, E. S. (1973). Tests for Departure from Normality. Empirical Results for the Distributions of b2 and √ b1. *Biometrika*, *60*(3), 613–622. JSTOR. https://doi.org/10.2307/2335012

Dupre, D., Akpan, D., Elias, E., Adam, J.-M., Meillon, B., Bonnefond, N., Dubois, M., & Tcherkassof, A. (2015). Oudjat: A configurable and usable annotation tool for the study of facial expressions of emotion. *International Journal of Human-Computer Studies*, *83*, 51–61. https://doi.org/10.1016/j.ijhcs.2015.05.010

Felder, R. M., & Brent, R. (2003). Designing and Teaching Courses to Satisfy the ABET

Engineering Criteria. *Journal of Engineering Education*, *92*(1), 7–25. https://doi.org/10.1002/j.2168-9830.2003.tb00734.x

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring Participation among Elementary Grade Students. *Educational and Psychological Measurement*, *51*(2), 393–402. https://doi.org/10.1177/0013164491512013

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and Inattentive-Withdrawn Behavior and Achievement among Fourth Graders. *The Elementary School Journal*, *95*(5), 421–434. https://doi.org/10.1086/461853

Finn, J., & Zimmer, K. (2012). Student Engagement: What Is It? Why Does It Matter? In *Handbook of research on student engagement* (pp. 97–131). https://doi.org/10.1007/978-1-4614-2018-7_5

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109. JSTOR.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763–782). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_37

Garcia, T., & Pintrich, P. R. (1996). Assessing Students' Motivation and Learning Strategies in the Classroom Context: The Motivated Strategies for Learning Questionnaire. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 319–339). Springer Netherlands. https://doi.org/10.1007/978-94-011-0657-3_12

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Gonzalez, J., & Wagenaar, R. (2003). *Tuning Educational Structures in Europe. Final Report – Pilot Project Phase 1*. University of Deusto Press. https://www.rug.nl/research/portal/publications/tuning-educational-structures-in-europe-final-report--pilot-project-phase-1(95d4ab16-6ba3-4324-a865-a87b765fa27e).html

Hale, A. R. (1984). Is safety training worthwhile? *Journal of Occupational Accidents*, *6*(1), 17–33. https://doi.org/10.1016/0376-6349(84)90026-9

Halko, N., Martinsson, P. G., & Tropp, J. A. (2009, September). *Finding Structure with Randomness: Stochastic Algorithms for Constructing Approximate matrix Decompositions* (Report or Paper No. 2009–05). California Institute of Technology. https://resolver.caltech.edu/CaltechAUTHORS:20111012-111324407

Haslam, R. A., Hide, S. A., Gibb, A. G. F., Gyi, D. E., Pavitt, T., Atkinson, S., & Duff, A. R. (2005). Contributing factors in construction accidents. *Applied Ergonomics*, *36*(4), 401–415. https://doi.org/10.1016/j.apergo.2004.12.002

Heckhausen, H. (1991). Achievement Motivation. In H. Heckhausen (Ed.), *Motivation and Action* (pp. 199–242). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-75961-1_8

Hinze Jimmie, & Gambatese John. (2003). Factors That Influence Safety Performance of Specialty Contractors. *Journal of Construction Engineering and Management*, *129*(2), 159–164. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:2(159)

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an Understanding of Definitions and Measures of School Engagement and Related Terms. *The California School Psychologist*, *8*(1), 7–27. https://doi.org/10.1007/BF03340893

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Levinson, S. C. (1998). Putting linguistics on a proper footing: Explorations in Goffman's participation framework. In P. Drew & A. Wooton (Eds.), *Erving Goffman: Exploring the Interaction Order* (pp. 161–227). Polity Press.

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

McCaslin, M., & Good, T. L. (1996). The informal curriculum. In *Handbook of educational psychology* (pp. 622–670). Prentice Hall International.

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*(3), 587–596. https://doi.org/10.1016/S0191-8869(03)00118-1

Mendez, G., Ochoa, X., Chiluiza, K., & De Wever, B. (2014). Curricular Design Analysis: A Data-Driven Perspective. *Journal of Learning Analytics*, *1*, 84–119. https://doi.org/10.18608/jla.2014.13.6

Namian Mostafa, Albert Alex, Zuluaga Carlos M., & Jaselskis Edward J. (2016). Improving Hazard-Recognition Performance and Safety Training Outcomes: Integrating Strategies for Training Transfer. *Journal of Construction Engineering and Management*, *142*(10), 04016048. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001160

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science Engineering*, *9*(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*, 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, *101*(1), 115–135. https://doi.org/10.1037/a0013383

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in Education* (pp. 13–36). Academic Press. https://doi.org/10.1016/B978-012372545-5/50003-4

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology*, *Complete*(42), 82–96. https://doi.org/10.1016/j.cedpsych.2015.05.002

Robotham, G. (2001). SAFETY TRAINING That Works. *Professional Safety*, *46*(5), 33.

Rumberger, R. W., & Lim, S. A. (2008). *Why Students Drop Out of School: A Review of 25 Years of Research* (No. 15; California Dropout Research Project, p. 130). University of California, Santa Barbara. http://cdrpsb.org/pubs_reports.htm

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with

Python. *Proceedings of the 9th Python in Science Conference*, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Skinner, E., & Belmont, M. (1993). Motivation in the classroom: Reciprocal effect of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581. https://doi.org/10.1037/0022-0663.85.4.571

Sousa, V., Almeida, N. M., & Dias, L. A. (2015). Risk-based management of occupational safety and health in the construction industry – Part 2: Quantitative model. *Safety Science*, *74*, 184–194. https://doi.org/10.1016/j.ssci.2015.01.003

Sparer, E. H. (2015). *Improving Health and Safety in Construction: The Intersection of Programs and Policies, Work Organization, and Safety Climate* [Doctoral, Harvard T.H. Chan School of Public Health]. http://nrs.harvard.edu/urn-3:HUL.InstRepos:16121136

Taylor, E. L. (2015). Safety benefits of mandatory OSHA 10h training. *Safety Science*, *77*, 66–71. https://doi.org/10.1016/j.ssci.2015.03.003

Tobon, S. (2007). El enfoque complejo de las competencias y el diseño curricular por ciclos propedéuticos. *Acción Pedagógica*, *16*(1), 14–28.

Turner, J. C., & Meyer, D. K. (2000). *Studying and Understanding the Instructional Contexts of Classrooms: Using our Past to Forge our Future*. https://doi.org/10.1207/s15326985ep3502_2

Van Zelst, R. H. (1954). The Effect of Age and Experience upon Accident Rate. *The Journal of Applied Psychology*, *38*(5), 313–317.

Waehrer, G. M., & Miller, T. R. (2009). *Does Safety Training Reduce Work Injury in the United States*. https://doi.org/10.2174/1875934300902010026

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., &

Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432–445. https://doi.org/10.1002/pits.20307

Wilkins, J. R. (2011). Construction workers' perceptions of health and safety training programmes. *Construction Management and Economics*, *29*(10), 1017–1026. https://doi.org/10.1080/01446193.2011.633538